50 Years since GW66

J A Greenwood   Cambridge

Emeritus Reader in Tribology

The GW theory really is 52 years old!

Proceedings of the

2nd International Symposium on Electric Contact Phenomena

GRAZ AUSTRIA

MAY 4,5,6 1964

Edited by P. A. KLAUDY

Technische Hochschule GRAZ

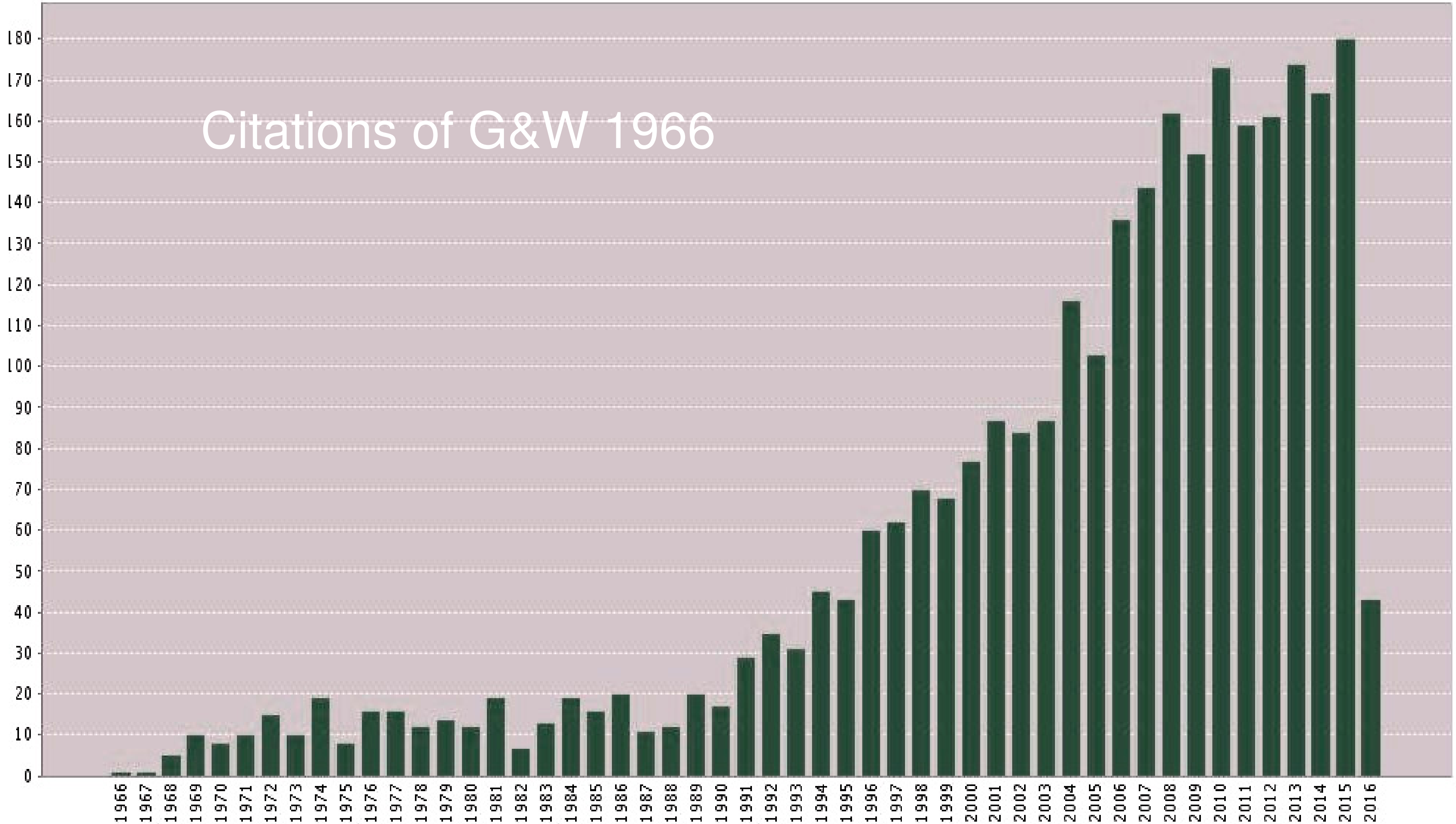
Lehrstuhl für Grundlagen der Elektrotechnik

Research Report No. 15
July 24, 1964

BURNDY
RESEARCH
DIVISION

The Contact of Nominally Flat Surfaces

J. A. Greenwood
J. B. P. Williamson



Total by April 2016 was 2959. But have they really read it?

Of course it all began with Archard……..
and this was long before the concept of a fractal *was invented.*

And perhaps we took over the idea of multiple Hertzian contacts from him

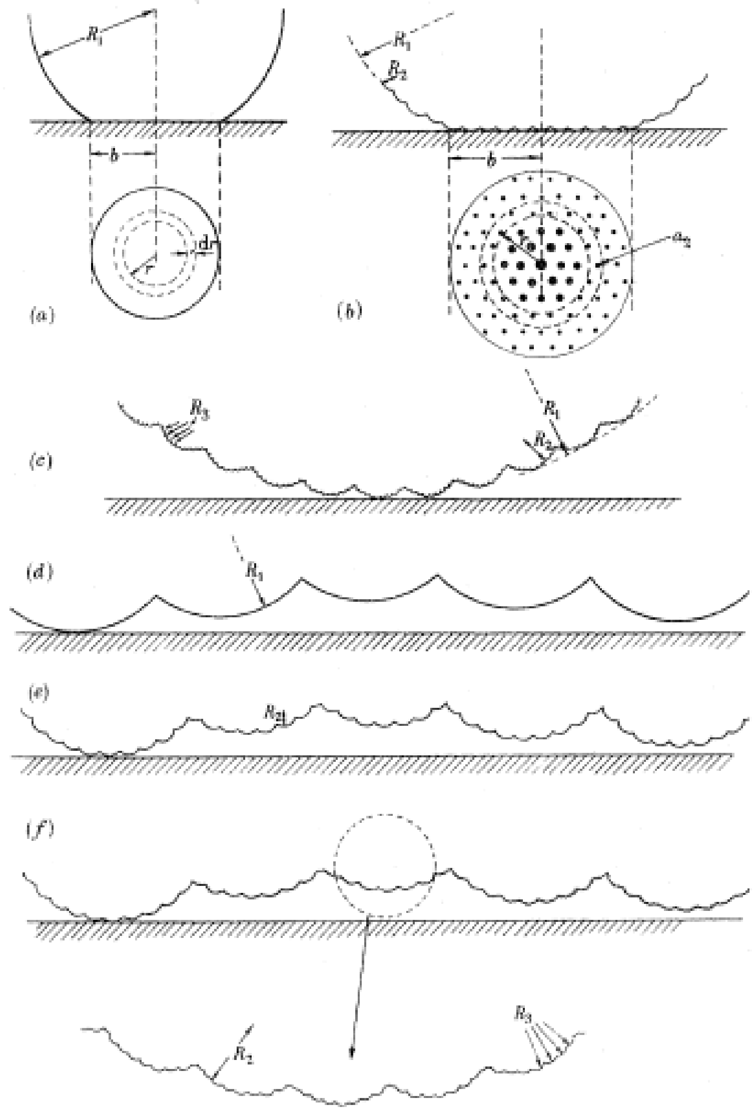


FIGURE 1. Models used in the theory. In (*a*) and (*b*) the surfaces are shown under a load, in (*c*)–(*f*) under zero load. The deduced relations between A and W for these models are (*a*) $A \propto W^{\frac{2}{3}}$, (*b*) $A \propto W^{\frac{8}{9}}$, (*c*) $A \propto W^{\frac{26}{27}}$, (*d*) $A \propto W^{\frac{4}{5}}$, (*e*) $A \propto W^{\frac{14}{15}}$, (*f*) $A \propto W^{\frac{44}{45}}$.

The same year (1958) F F Ling published his contact analysis, accompanied by some good experimental load ***v*** approach data.

So what did he do wrong, ..
or what did we do right, so that we, not Ling, became the standard reference?

Ling assumed, naturally, that the asperities deformed plastically. And he offered too many asperity shapes [*and an implausible fracture mechanism*], and too many possible height distributions ***without measuring any***

But what really mattered was his choosing the point of first contact as his datum, and seeking a power law relation between load and approach.

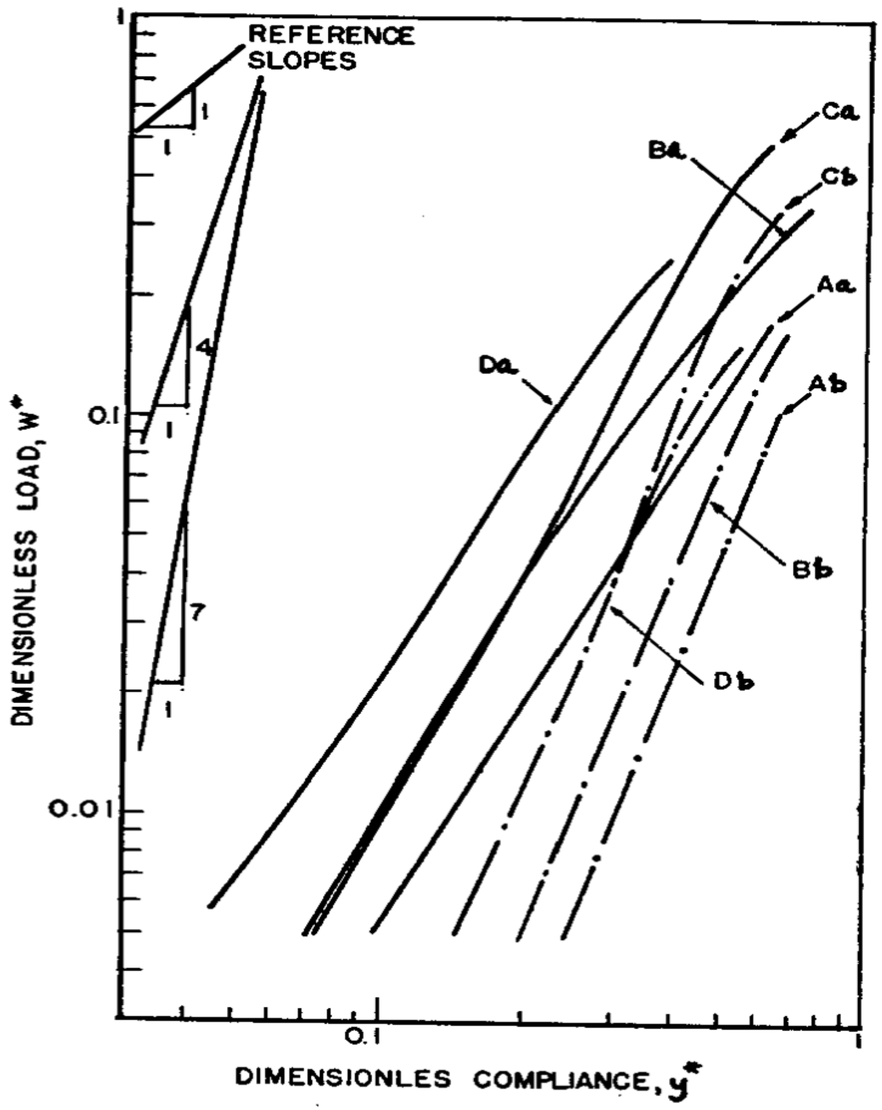


FIG. 1. Log-log plot of dimensionless load *vs* dimensionless compliance of eight selected models. Aa—compression by a rigid flat of a set of rigid, perfectly-plastic wedges or hemispheres uniformly distributed through a depth of unity from the surface. Ba—same distributed linearly. Ca—same distributed normally. Da—same having a Poisson distribution. Ab, Bb, Cb, and Db—same as Aa, Ba, Ca, and Da respectively except cones are compressed.

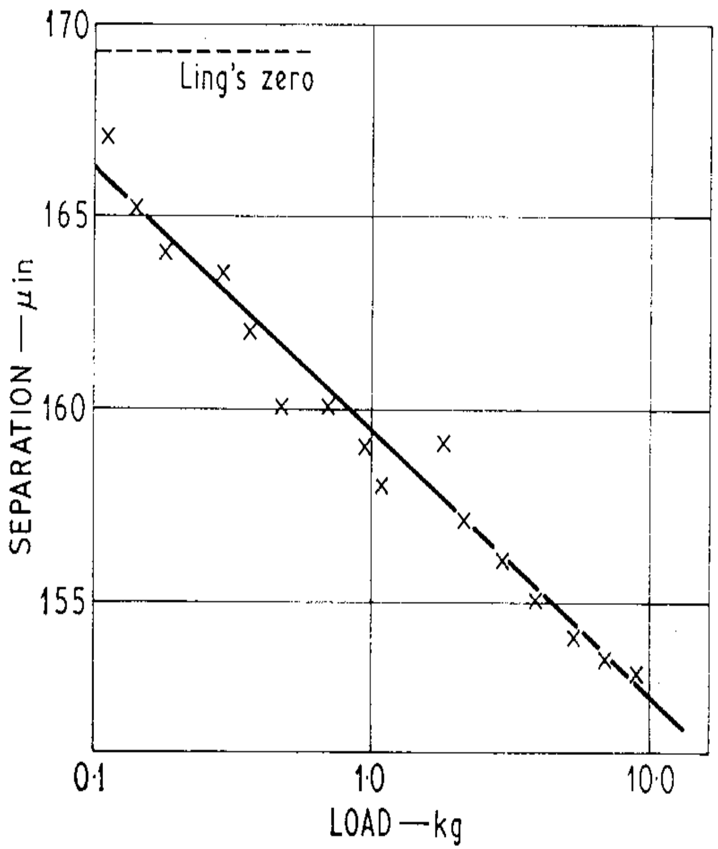


Fig. 7. Replot of Ling's experimental results (Fig. 9 of (15)) Ling's original criticism of single-rough-surface models was based on the discrepancy between experimental and theoretical curves on a log–log plot. On a log–linear plot, the discrepancy does not appear and it seems that it arises from an invalid extrapolation to zero load to find the origin for the log–log plot

Proc Instn Mech Engrs 1970–71

Ling plotted his data on a log-log plot W(δ) ..but this just gave a curve, with a slope increasing from 2 to 8

The point of first contact is an unreliable datum.
Even with a large population it will be erratic, and its neighbours can be ***anywhere***

In 1963 SKF didn't have a computer in the laboratory either!

Lubricant Films in Rolling Contact of Rough Surfaces

T. E. TALLIAN, Y. P. CHIU, D. F. HUTTENLOCHER, J. A.KAMENSHINE, L. B. SIBLEY, & N. E. SINDLINGER

Surface microgeometry of the rolling tracks on the balls is statistically analyzed by processing electrical analogs of surface profiles through on-line computing equipment.

The output of the surface tracing instrument was fed into an FM magnetic tape recorder according to diagram A, Fig. 10. The low-frequency band pass filter inserted between the surface tracing instrument and the tape recorder was set to a pass band of calculated width. It will be seen that this finite band width is necessary for usable results.

A. RECORDING

TRANSDUCER	⇒	CARRIER AMPLIFIER & DEMODULATOR	⇒	0.02 – 20 CPS BAND PASS FILTER	⇒	AMPLIFIER	⇒	F M TAPE RECORDER

The tape-recorded electrical analog of the surface profile was then processed through the circuitry shown in block diagram B of Fig. 10. The arrangement has a common input consisting of the playback system of the FM tape recorder, an amplifier, and a variable band pass filter which, again, was set as explained later

-- -- -- -- -- -- -- -- -- -- --

Output Channel 1 is a level discriminator, operating in conjunction with a 400 channel memory, being swept by an internal clock at a predetermined rate. [*To give a frequency distribution of dwell times at a chosen level*]

-- -- -- -- -- -- -- -- -- -- --

Output channel 2 comprises circuitry to obtain an amplitude histogram of the signal by way of periodic sampling triggered by a pulse generator. Each momentary amplitude sampled is converted in the amplitude time converter to a proportional time interval. The time intervals are used to accumulate counts in the memory unit as described for Channel 1, giving the amplitude distribution. Print-out is performed on command.

They were the first *tribologists* to think of Gaussian height distributions and to use signal theory to understand contacts

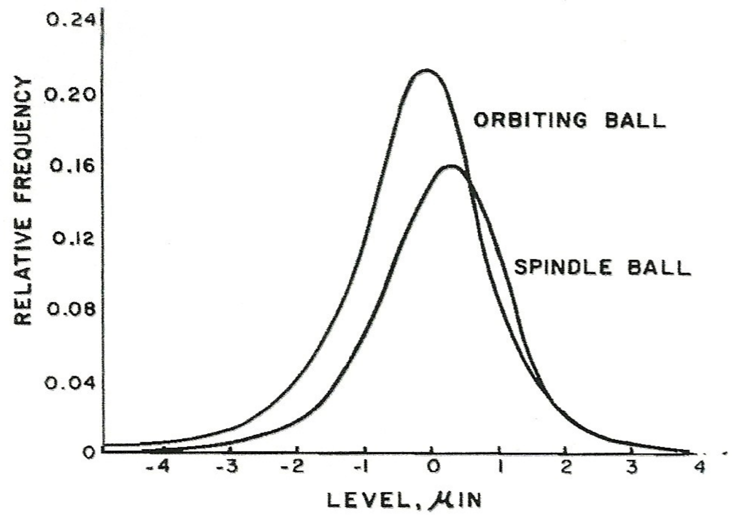


FIG. 11A. Roughness amplitude distribution; frequency diagram on linear scale.

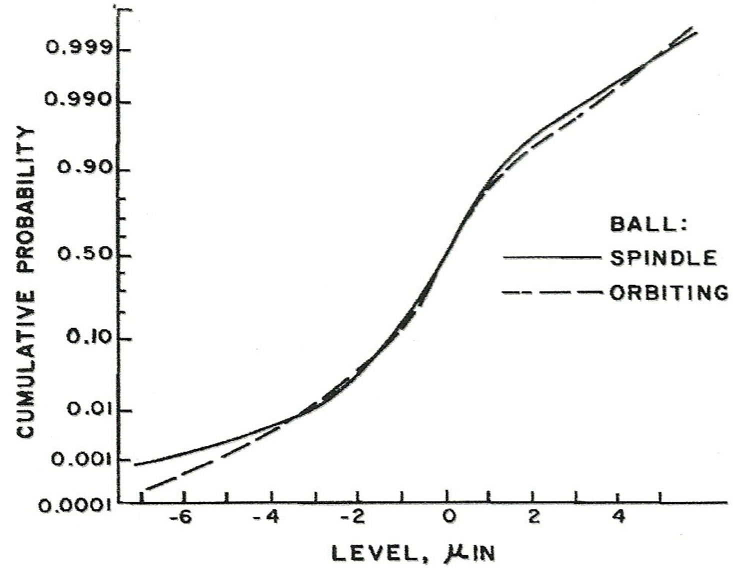


FIG. 11B. Roughness amplitude distribution; cumulative distribution on Gaussian scale.

So why G&W 1966 ?

We actually measured surface roughness…but so did Abbott & Firestone in 1932, (and invented the bearing area curve): Bickell (1963) published work showing heights were gaussian ...by drawing lines on the pen recorder output and counting …Tallian's group found the height distribution by sorting the signal into a 400 channel memory….so was it ***feeding it into a computer*** that made the difference?. Or was it the (obvious) next step; using the computer to locate peaks, so we could plot their heights and curvatures …and link up with Archard's ideas?

Perhaps we just got the timing right: for the metrologists (Reason at Rank Taylor-Hobson; Sharman at the National Engineering Laboratory) had also begun to feed their signals into a computer.

But perhaps we got the statistical theory right, by focussing on means and standard deviations, and having nothing to do with extreme values? Or even by firmly avoiding the term *normal* distribution, and using the magic password ***Gaussian***?

Worn surfaces do not have Gaussian height distributions: and random field theory is inapplicable! But the higher peaks may well behave as Gaussian….

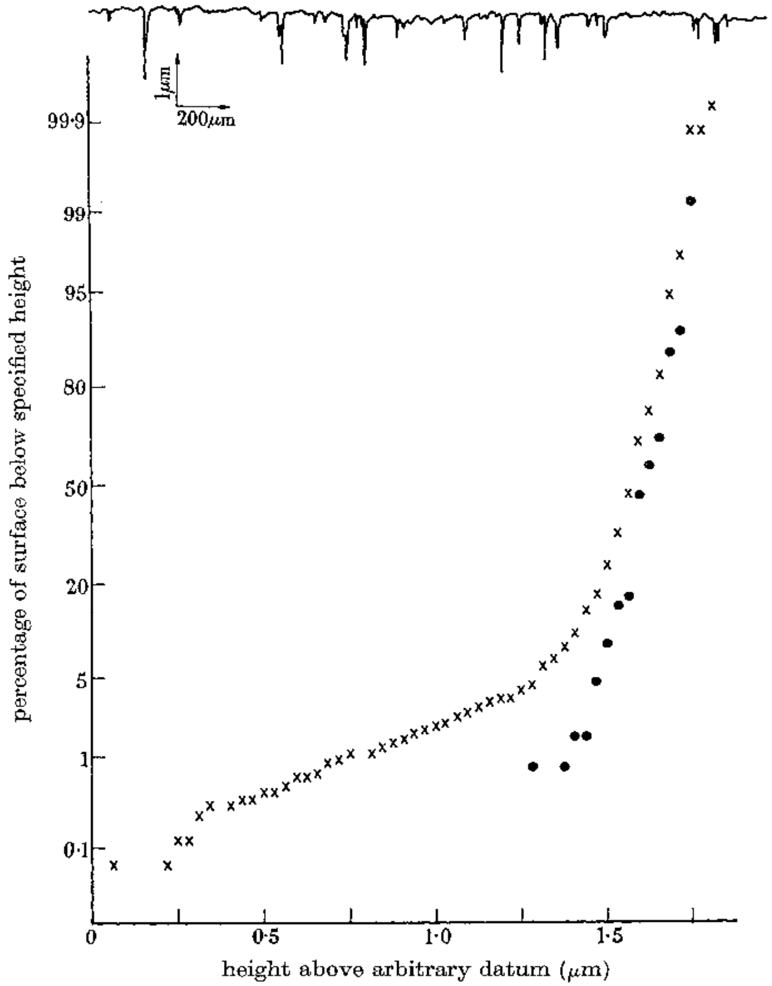
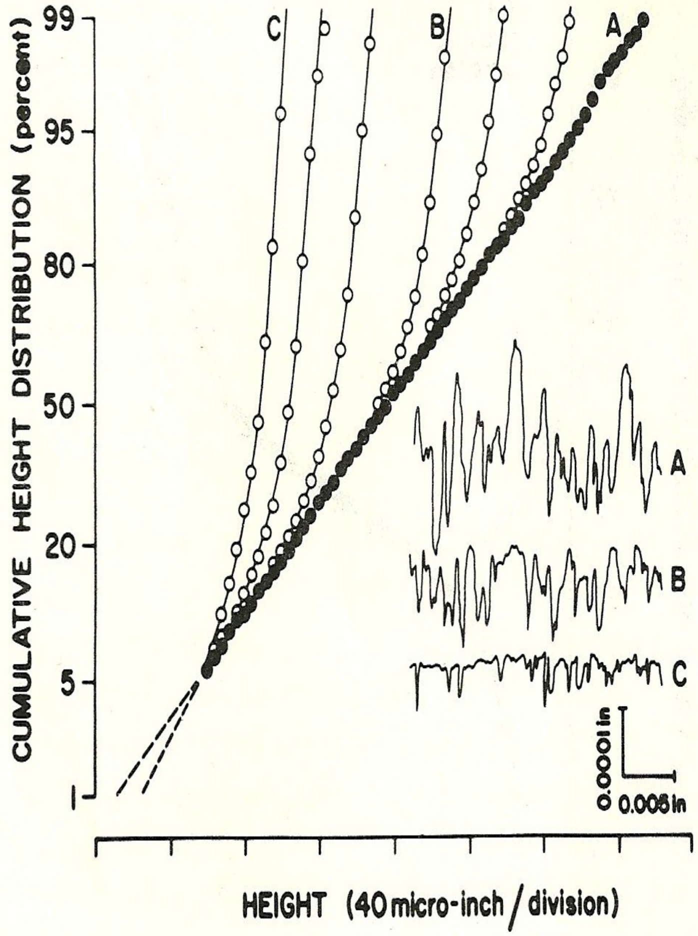


FIGURE 6. Cumulative height distribution of mild steel specimen. Distribution of all heights, ×. Distribution of peaks, ●. This specimen was abraded on 400 grade carborundum paper, then slid against a copper block flooded with oleic acid, at approximately 10 Kg, 130 cm/s for 30 s. Although the distribution is at first sight highly non-Gaussian, in fact nearly 90 % of the surface is approximately Gaussian; the surface, with an actual standard deviation of 1·3μm, would behave in contact as if Gaussian with a standard deviation of half this. The profile of the same surface is shown in the upper diagram: the vertical magnification is 200 times the horizontal magnification.

Probability paper is a more informative way of studying wear *than measuring skewness or kurtosis*



CUMULATIVE HEIGHT DISTRIBUTIONS SHOWING THE EFFECT OF WEAR ON THE INITIALLY GAUSSIAN HEIGHT DISTRIBUTION (●) OF A BEAD-BLASTED SURFACE. The six non-Gaussian distributions (O) represent, from right to left, progressive stages in the wearing process. The curves are related to each other by setting the heights of the fifth percentiles equal. The experimental points are omitted wherever

The GW theory.

For a single contact, the Hertz equations are

$$A_i = \pi R w; \qquad P_i = \tfrac{4}{3} E^* R^{1/2} w^{3/2} \qquad \text{where} \qquad \frac{1}{E^*} = \frac{1-\nu_1^2}{E_1} + \frac{1-\nu_2^2}{E_2} \ldots$$

{ so that if one body is rigid, E^* is the plane **strain** modulus of the other}

If the probability of a summit with height between z and $z+dz$ is $\phi(z)\,dz$, (and there are N summits), then when the separation of the mean planes is d, the number of contacts will be $n = N\int_{z=d}^{\infty} \phi(z)\,dz$,

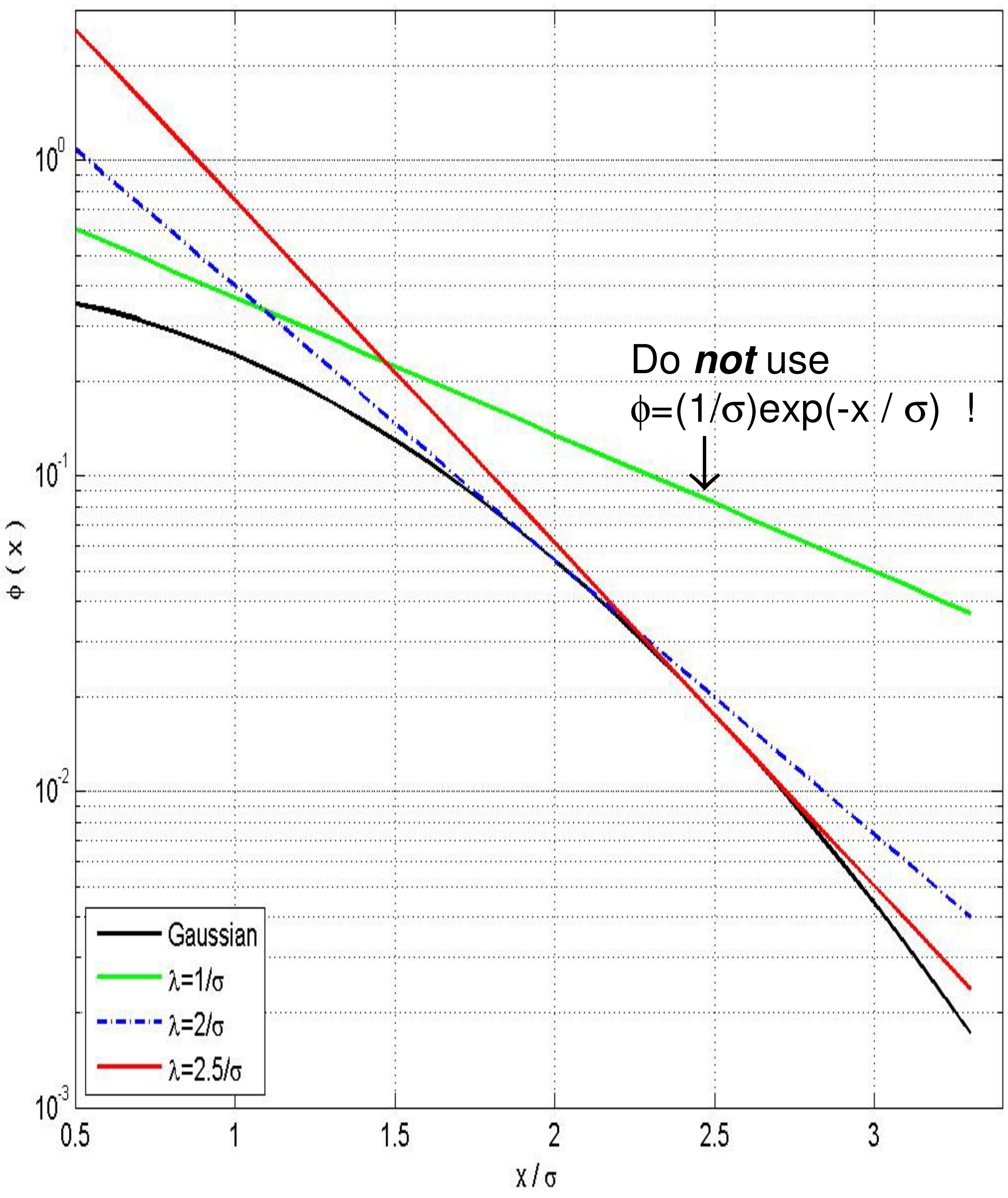
and the contact area and load will be

$$A = \pi NR \int_{z=d}^{\infty} (z-d)\,\phi(z)\,dz \;; \qquad P = \tfrac{4}{3} N E^* R^{1/2} \int_{z=d}^{\infty} (z-d)^{3/2}\,\phi(z)\,dz$$

[So if we use non-dimensional variables $h \equiv d/\sigma$; $\xi \equiv z/\sigma$, the real contact pressure will be $\frac{4}{3\pi} E^* \frac{\sigma^{1/2}}{R^{1/2}} \times$ ratio of the two integrals …..and the ratio varies only slowly with height]

Approximating a Gaussian by an exponential makes simple analysis possible: but do use the best exponential !

And since the skewness of an exponential is large while a gaussian has none, this should stop all investigations into the effect of skewness on contact behaviour !



To get the best approximation, fit at two points

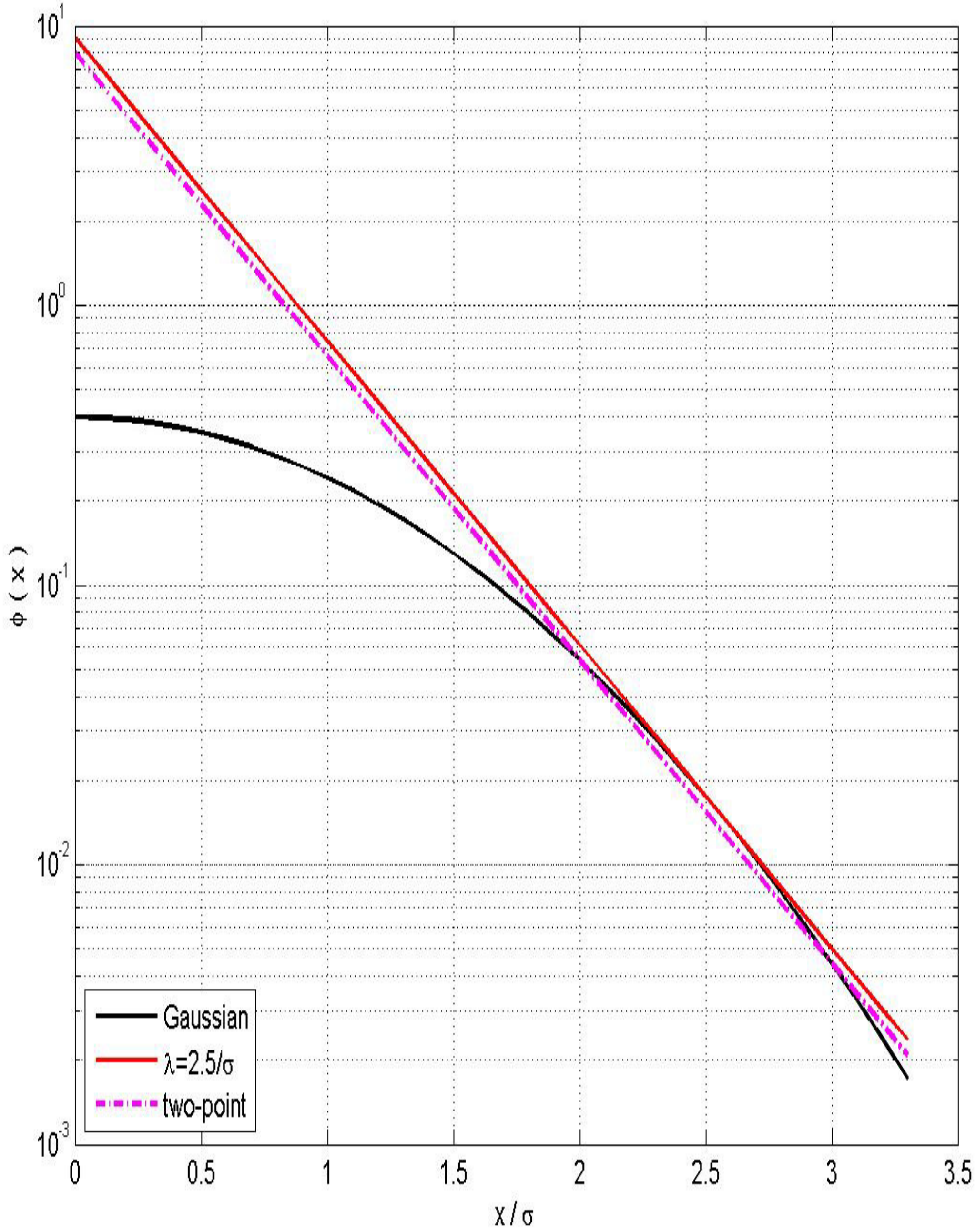
$\xi(1)$, $\xi(2)$:

then use

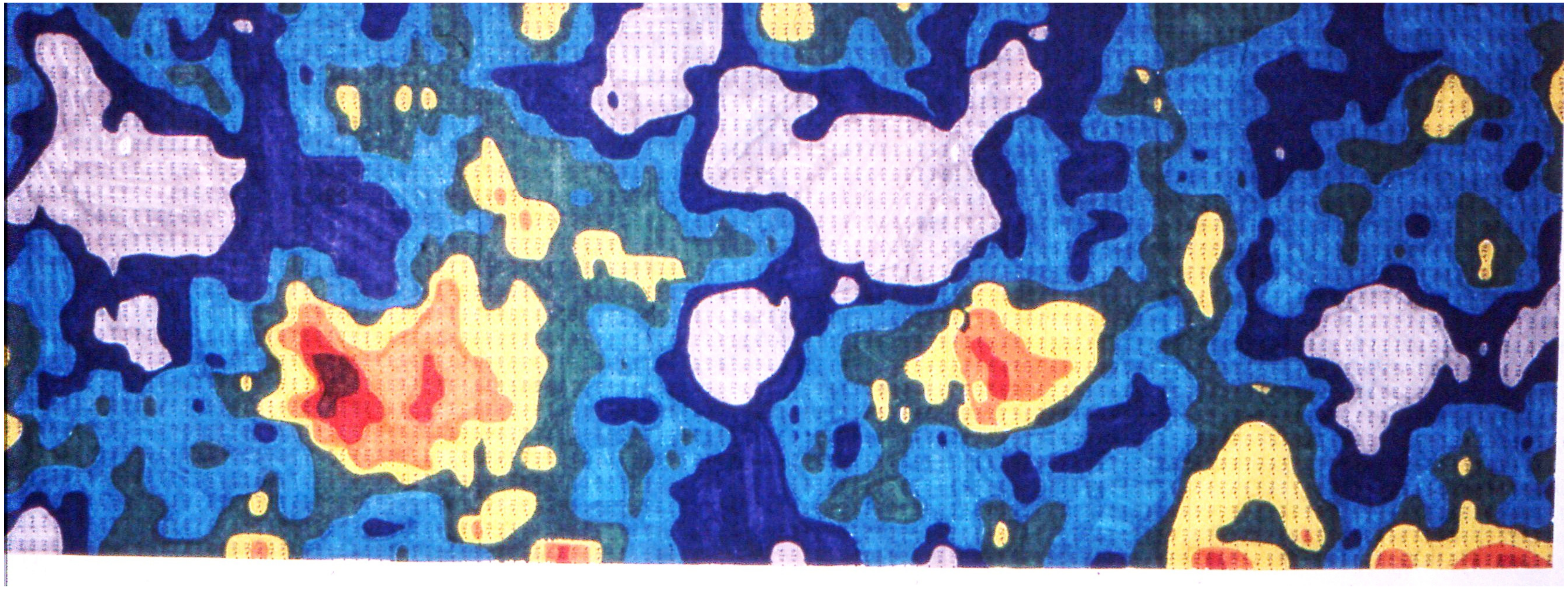
$\Phi(\xi)=\lambda \exp(-\lambda\xi)$

with

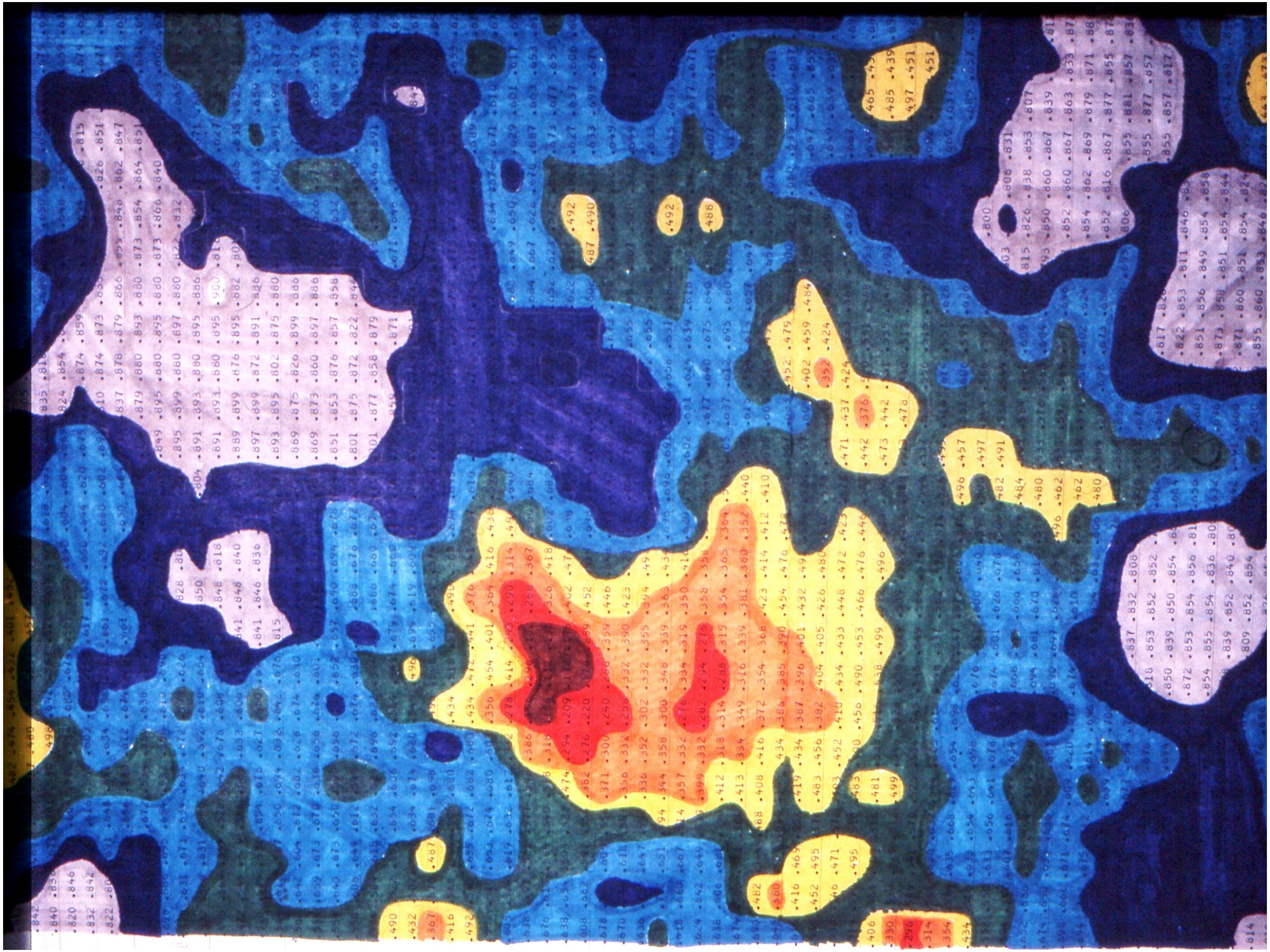
$\lambda=(1/2)[\xi(1)+\xi(2)]$

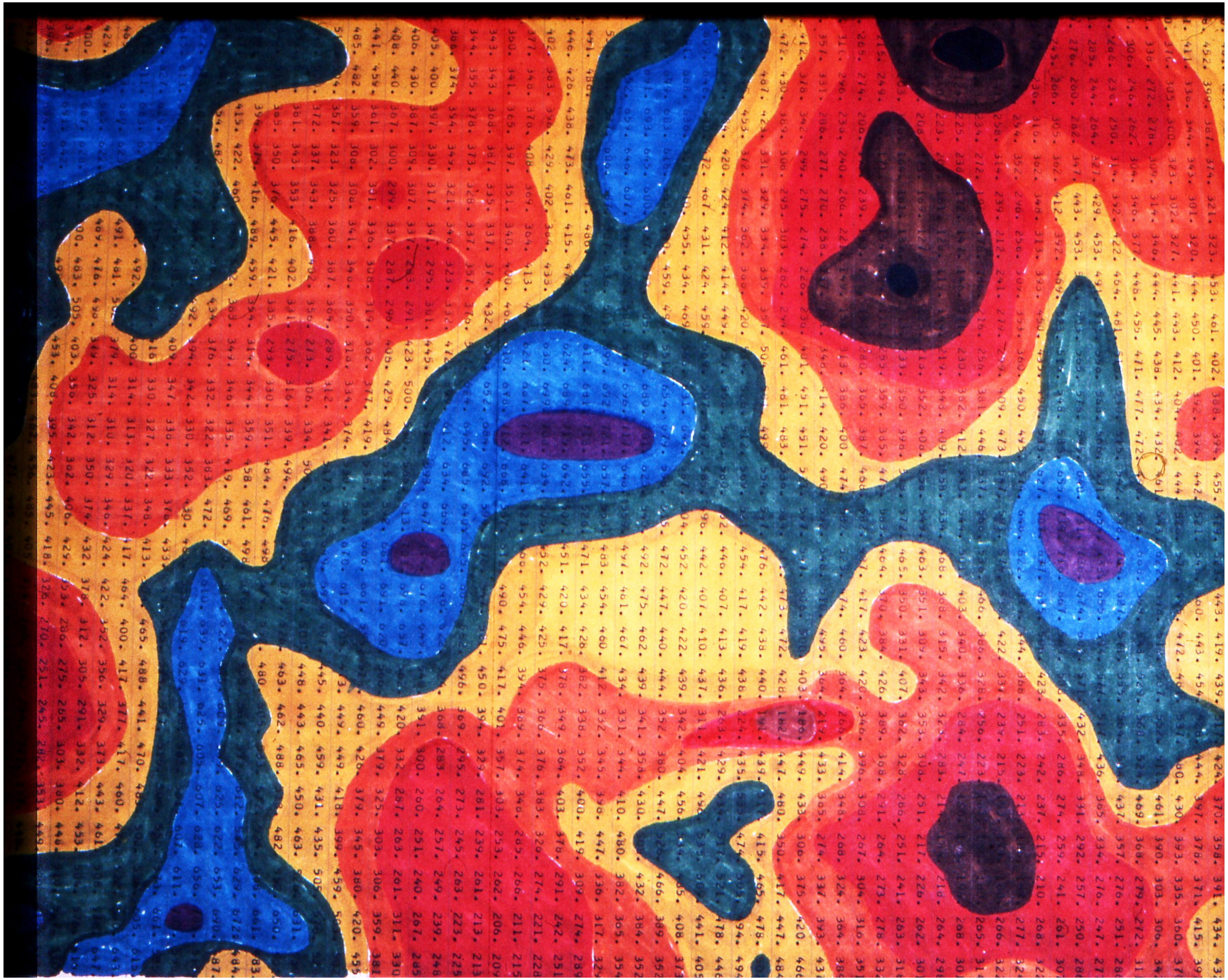


Glass blasted with alumina



Once you have a digitised profile, you can make repeated traversals, and assemble the information into a map. And then you learn how naïve it was to believe that a **peak** corresponds to an **asperity**…or even how the *number* of peaks might be used to estimate the *number* of asperities.





Glass-bead blasted aluminium. [sampling interval 1.7μm]

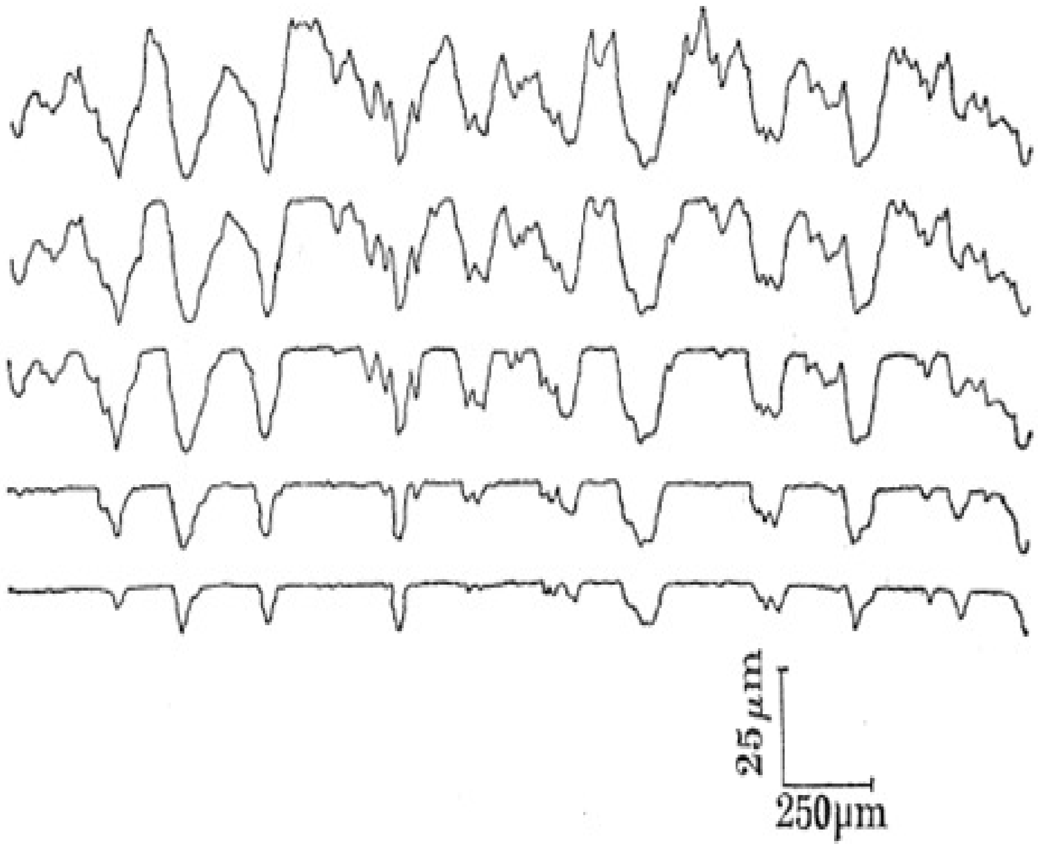
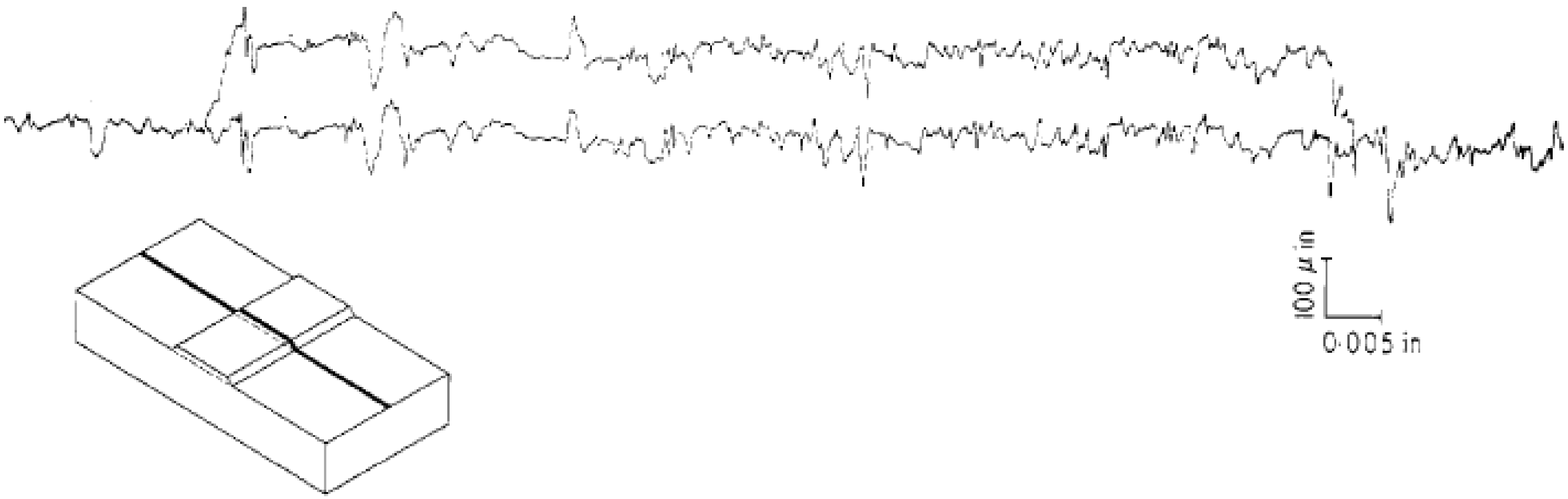


FIGURE 2. Profiles showing progressive deformation of asperities on aluminium specimen. From top to bottom: virgin surface, 1800, 9000, 34 000 and 54 000 lbf/in² nominal pressure. (1 lbf/in² ≈ 6.9 kPa.)

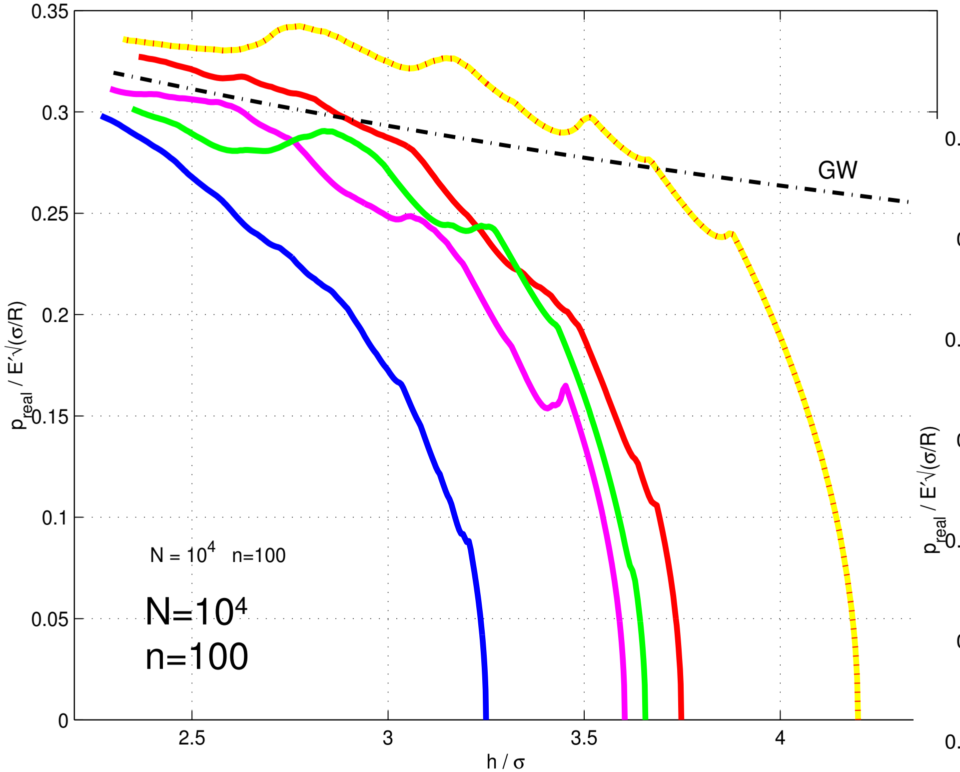
Figure 7. Electrodeposition of gold on copper.

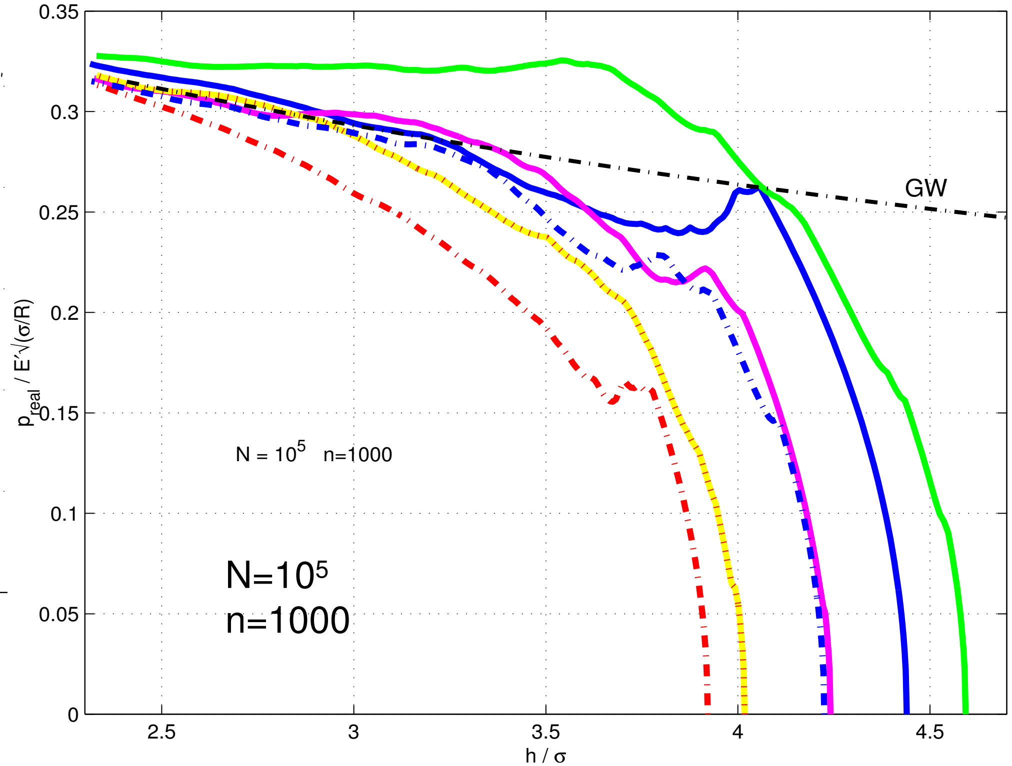
Profile of copper substrate with profile of gold deposit superposed, showing how the surface features of the copper persist in the new surface, even though the plate thickness is several times the height of a typical asperity. The inset shows the location of the profiles on the specimen before and after plating

Williamson & Hunt



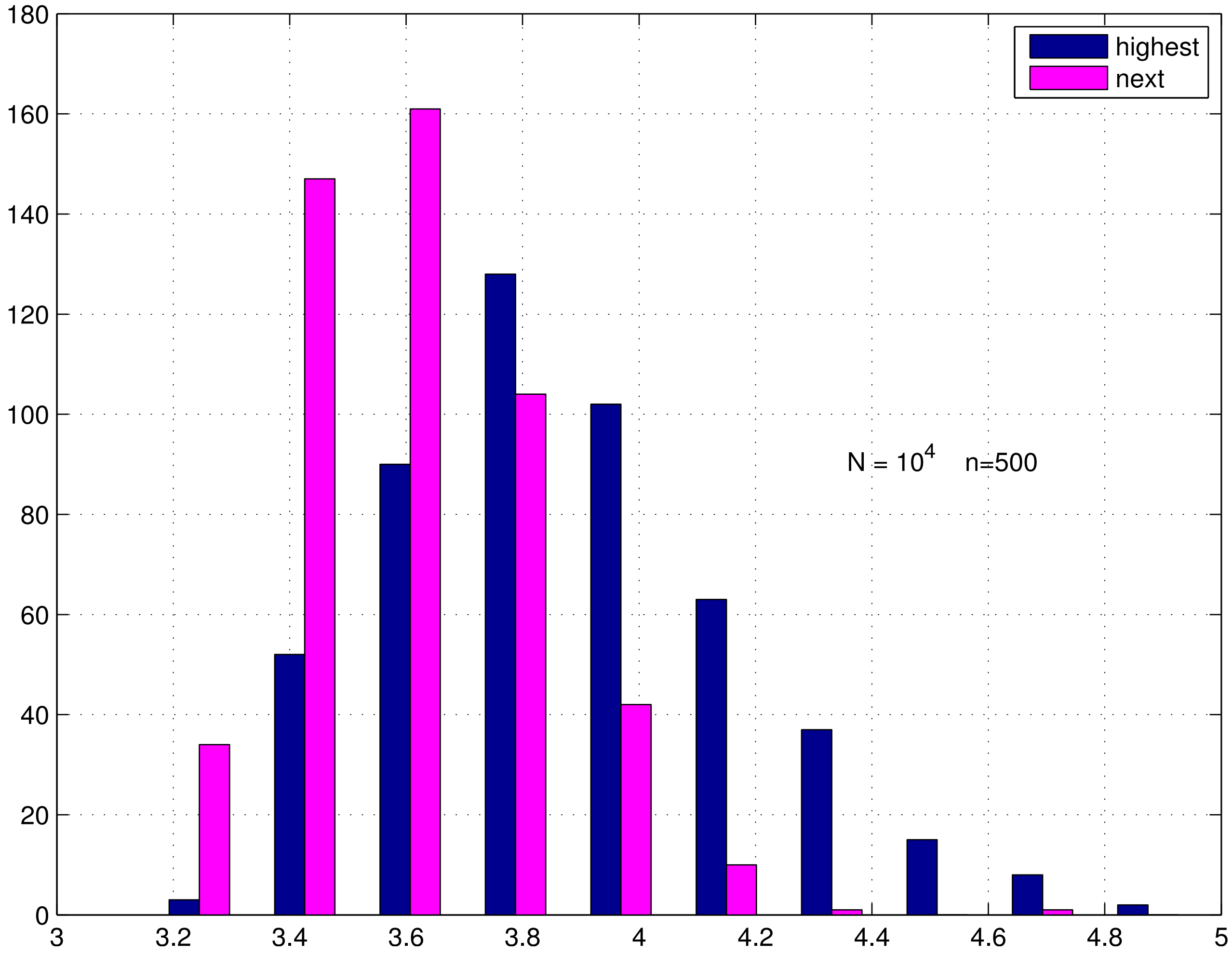
The approach for first contact varies enormously !
And with 100 contacts the scatter is still great.





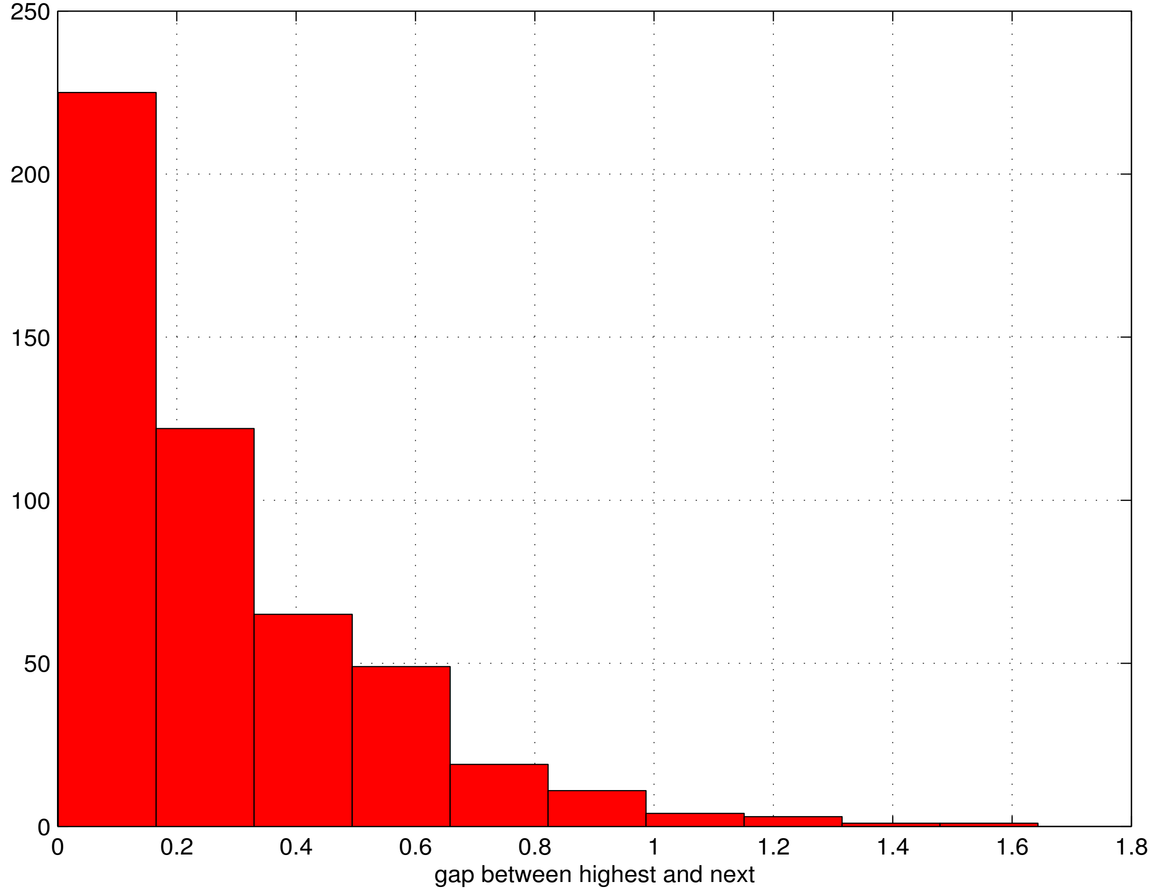
But ultimately the contact pressure begins to settle down…(to the proper value !)

Earlier I argued that Ling was misguided to base his analysis on the point of first contact: but recently I was persuaded by O'Shea to study it. So I generated 10^4 Gaussian heights, 500 times, and investigated where contact will occur:



And particularly, how big the gap between the first and second contacts will be:

often a whole rms height σ.

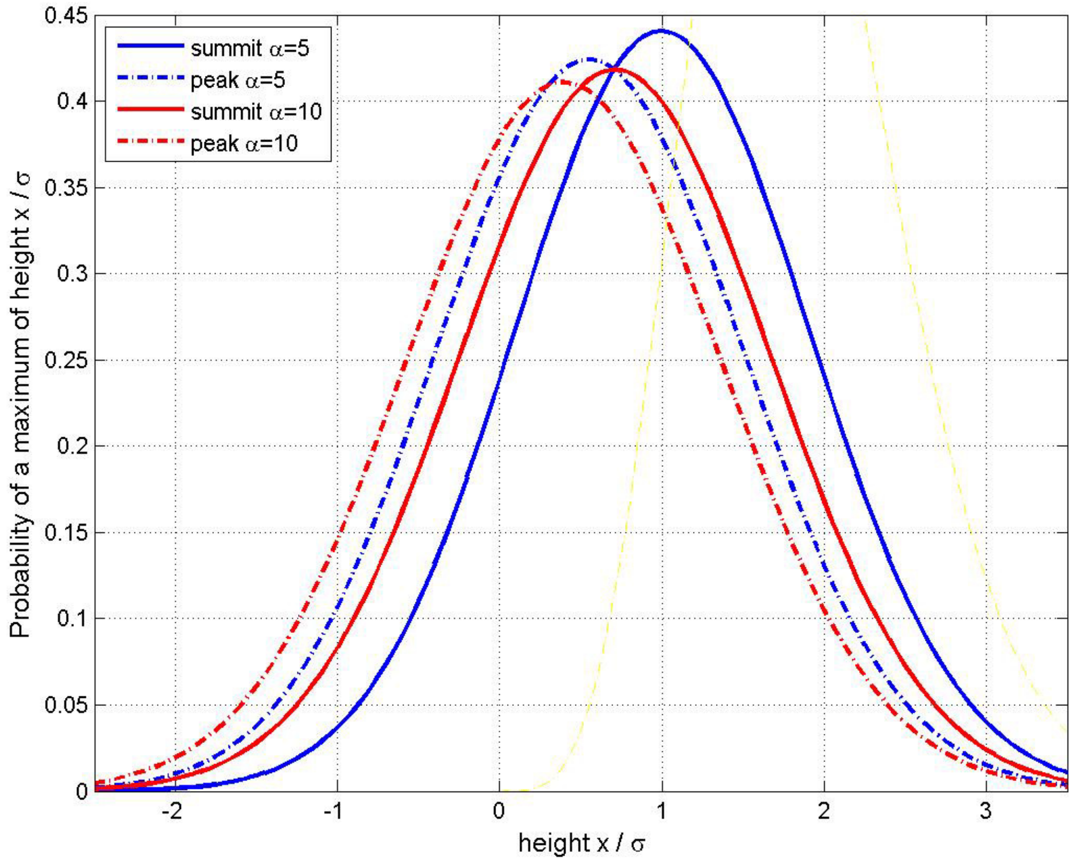


A valuable advance was Nayak's theory that surface roughness should be treated as a random field. Peaks are ***not*** summits..as is immediately obvious from a map: most peaks are ***shoulders,*** and do not even correspond to a summit. For a random field, we expect, summit density ≈ 1.2 (peak density)^2

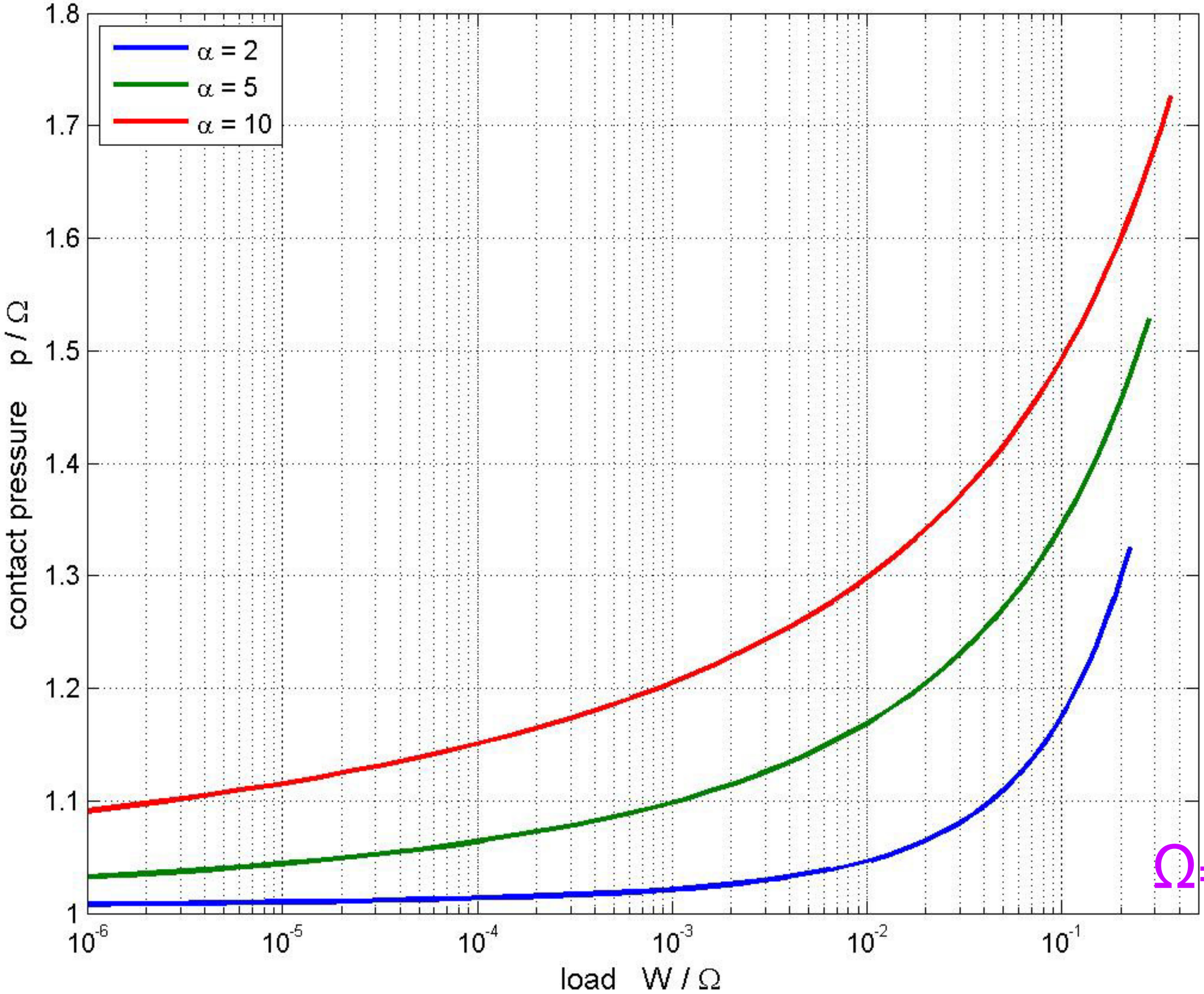
Nayak's Gaussian random field theory brought out the different properties of peaks and summits

…and predicted mean summit heights and curvatures from three easily measured quantities: the moments of the spectral density: m_0, m_2, m_4

$$\alpha = m_0 m_4 / m_2^2$$

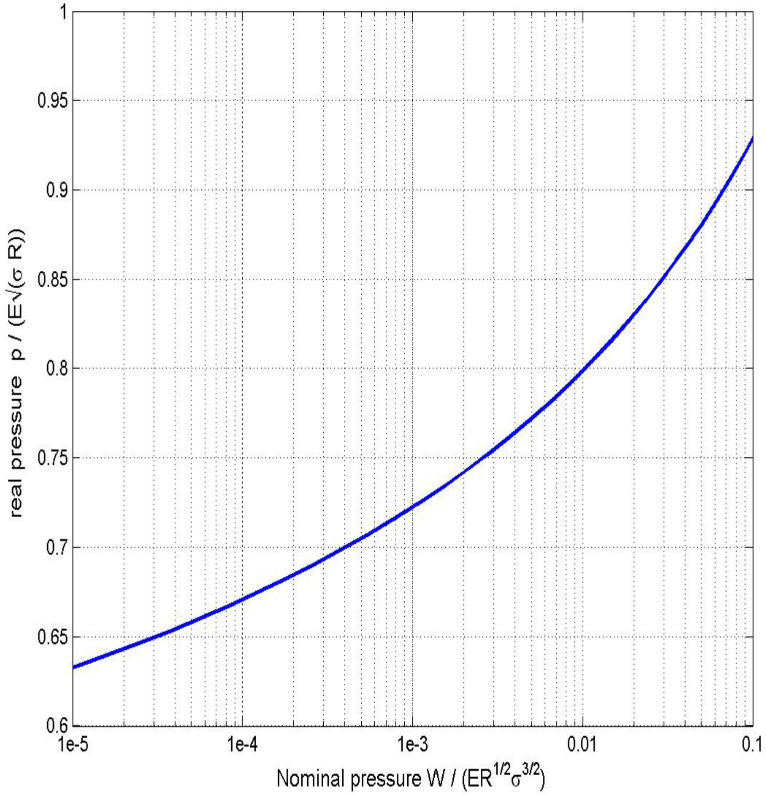


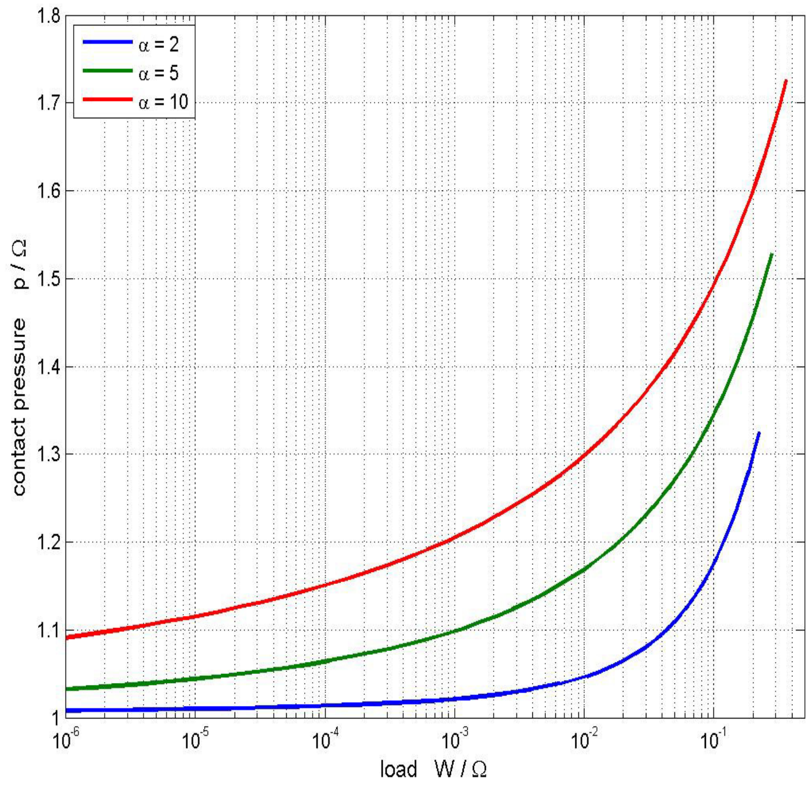
Bush, Gibson & Thomas used Nayak's summit and summit curvature distributions to do a full analysis of elastic contact of a Gaussian surface



$\Omega = E'\sqrt{(m_2/\pi)}$

The GW theory does not give proportionality between load and area, while the BGT theory does… ??

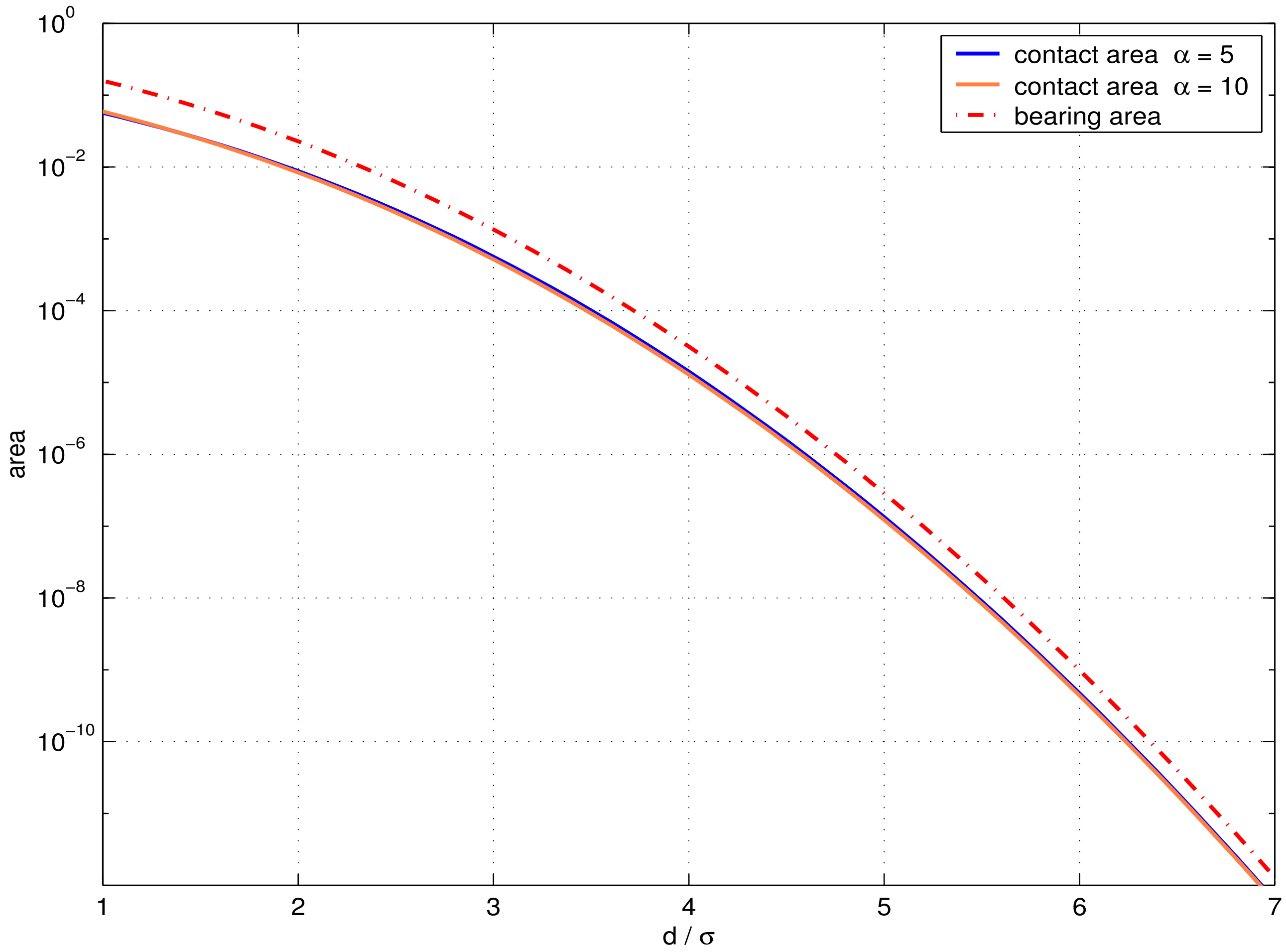




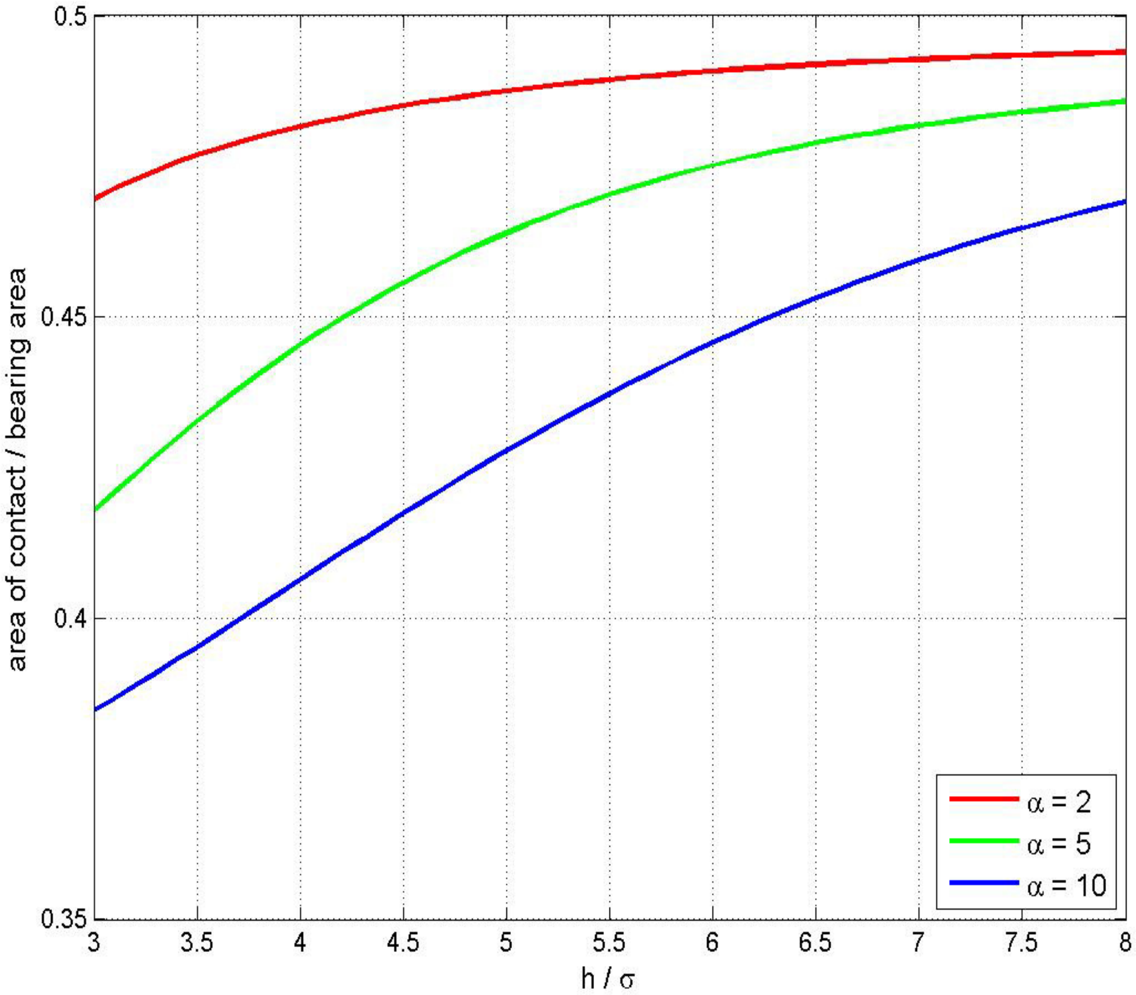
The vaunted asymptotic proportionality between contact area and load only occurs at impractically large separations and negligible loads

An astonishing prediction of the BGT analysis is the close correspondence between the contact area and the bearing area

How can two completely different quantities turn out to be so related?



It's not really very close: until you remember that both quantities vary over a range of 10^4



To understand the proportionality, we need to examine what Nayak's analysis says about the shape of asperities

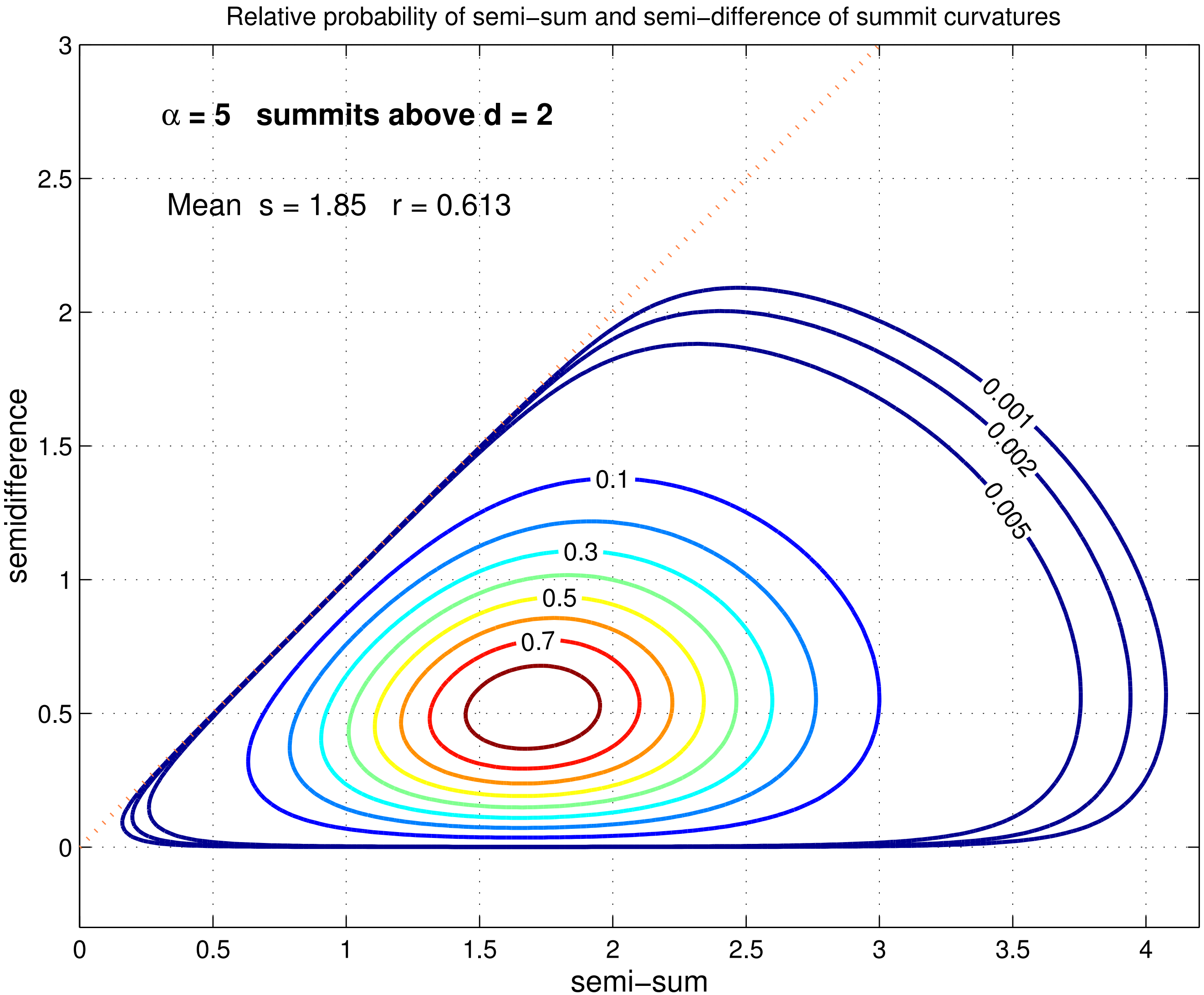
For Nayak himself ducks the issue, contenting himself with finding the mean summit curvature and showing that (and how) it increases with the summit height

Need we take into account the ellipticity of the summits?

Hertz theory for circular contacts, using $1/R=\sqrt{(\kappa_1\kappa_2)}$ is accurate to 0.1% for $\kappa_1/\kappa_2 < 2$. {and to 2% for $\kappa_1/\kappa_2 < 5$}

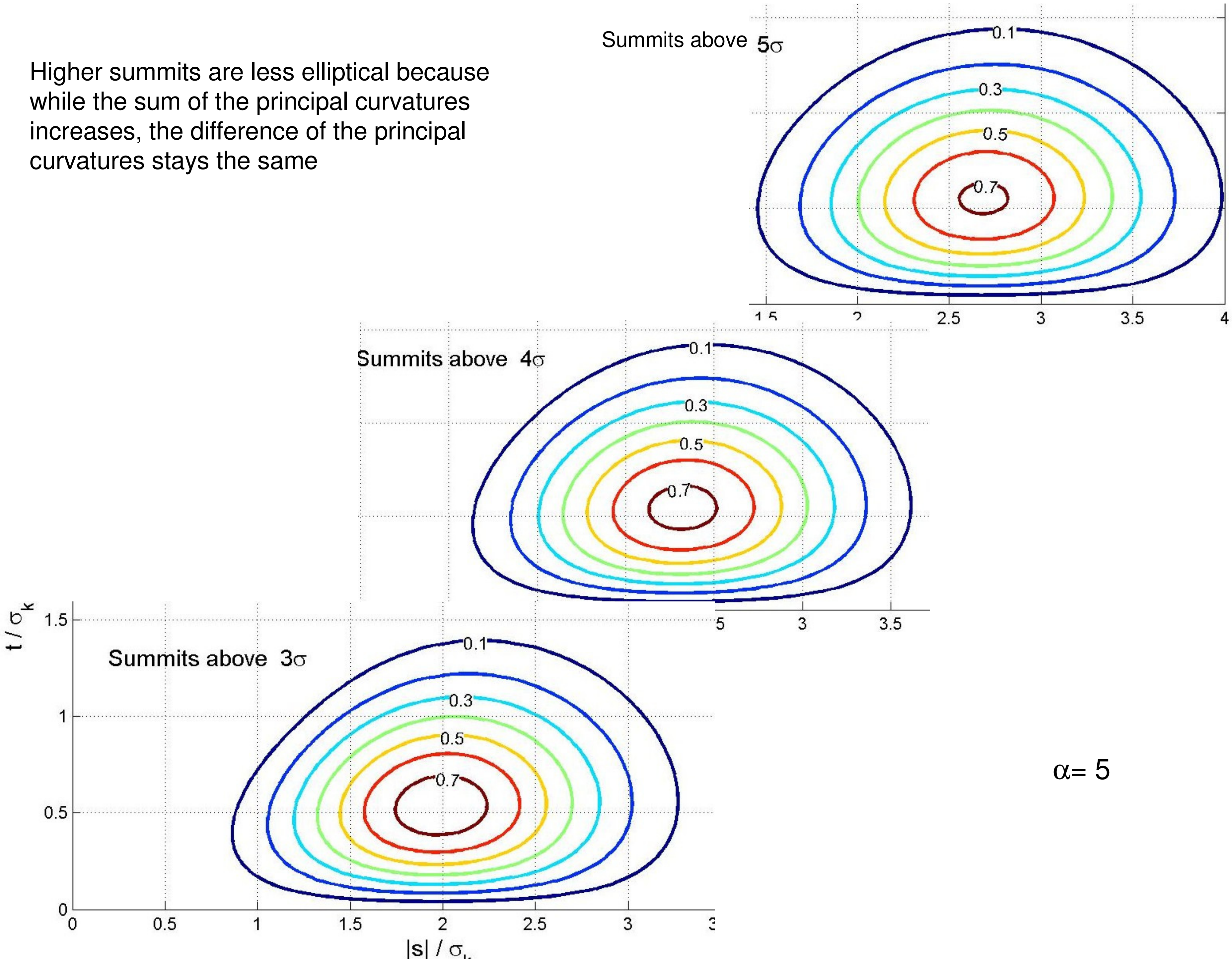
And for a circular Hertzian contact, the contact area is indeed exactly half the bearing area: $a^2=R\,\delta$ compared with $a^2=2\,R\,\delta$

Well, what shape are the summits?



There are ***no*** circular contacts **!**

Higher summits are less elliptical because while the sum of the principal curvatures increases, the difference of the principal curvatures stays the same



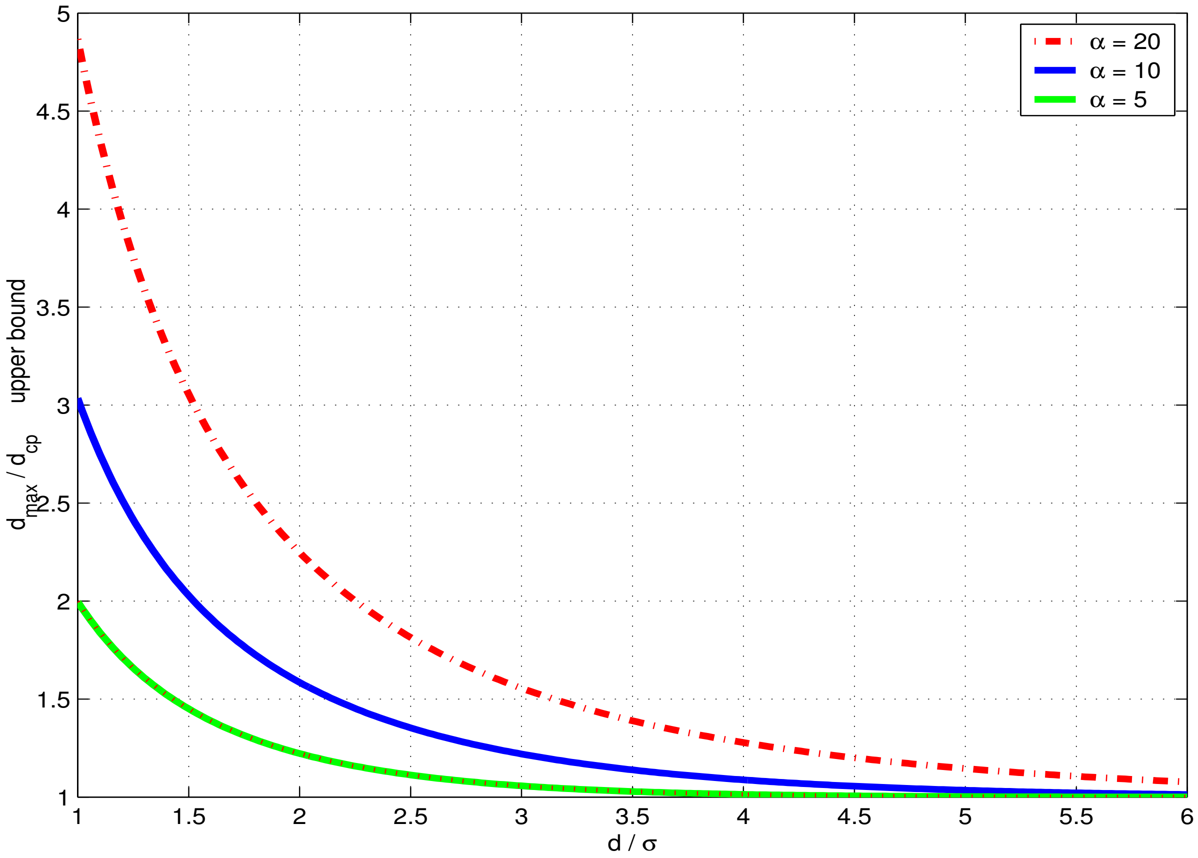
$\alpha = 5$

But of course Nayak's analysis tells us only about the summits of the asperities (as does GW peak-counting scheme); a group of circular summits may very quickly become a non-circular contact patch.

So we need to ask, are the summits isolated?

Nayak has the answer: he gives the number of contour areas at a given height: so we compare that number with the number of summits above that height

If each contour area contains only a single summit, the summits are isolated



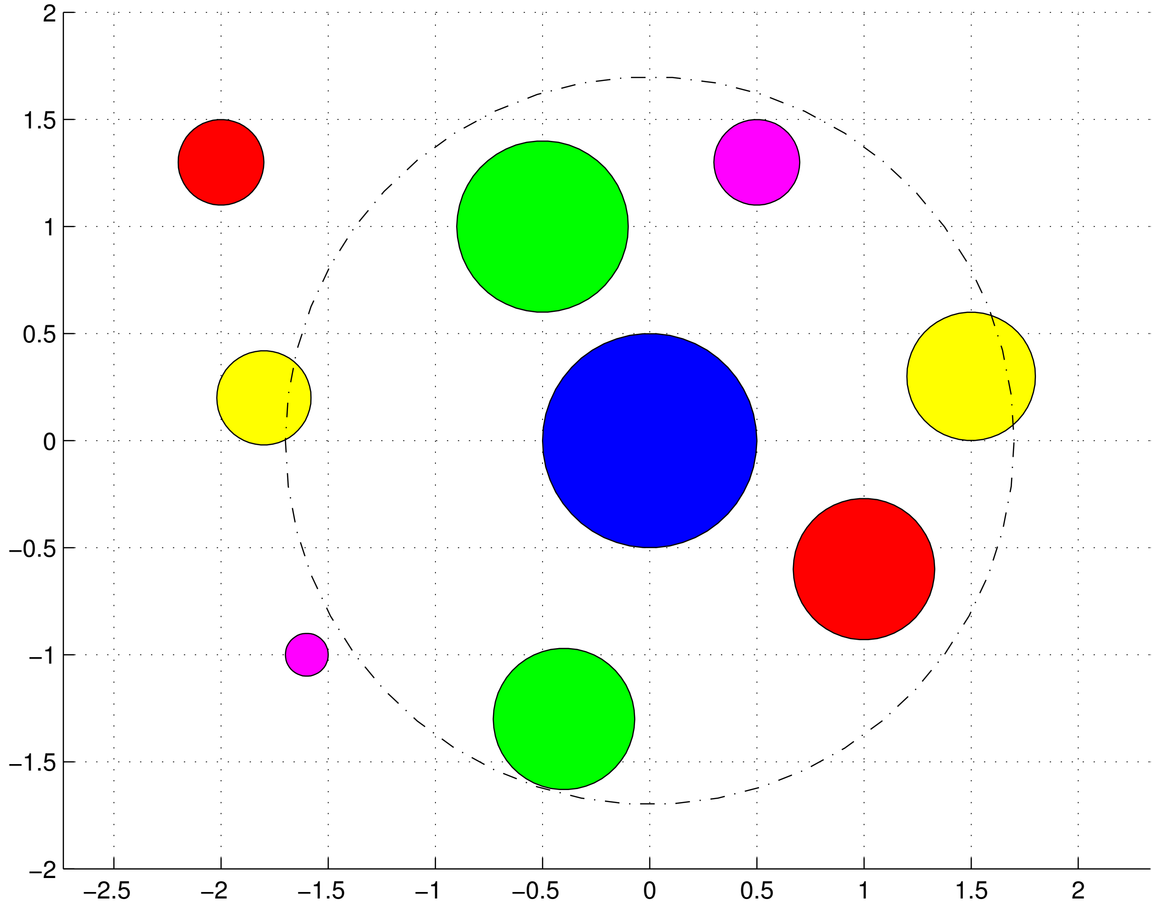
So yes, high summits are indeed isolated paraboloidal asperities

…and the contact area will indeed be just half of the bearing area

… if only we could find the load so easily !

Is the neglect of interaction by GW or BGT serious?

There is no doubt that a load on one contact spot will lower the neighbours, and may stop, or delay, them from making contact.



But is nearest neighbour interaction the real problem?

Olber's paradox: why is it dark at night?

If we live near just one of an infinite number of stars, with a density of *n* (per cubic light year?), then a spherical shell of radius R around us will contain $n.(\pi R^2\, dR)$ stars, each emitting light. But by the inverse square law, the illumination from a star at distance R will only be β/R^2: so the shell will contribute $n\beta/R^2.(\pi R^2 dR)$…ie $n\, \pi\beta.dR$.
So all the stars together give $n\, \pi\beta.\int dR$ ….. !

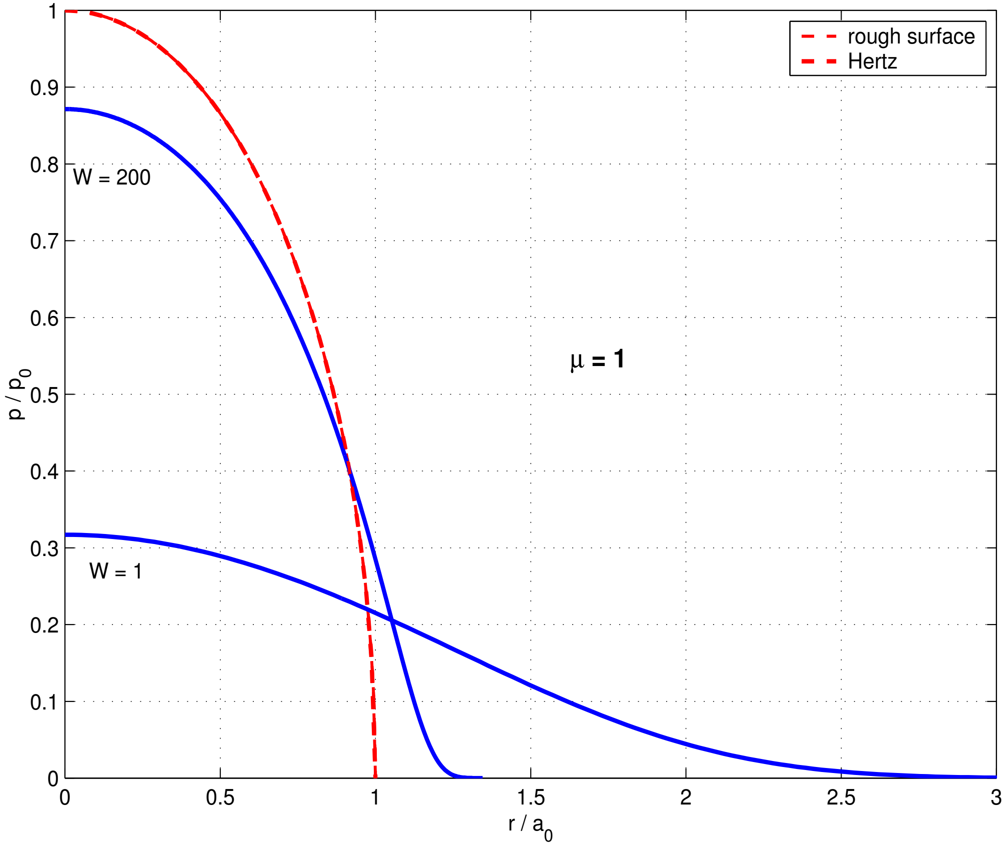
Need I go on?
A contact distant r will reduce the height by $P/\pi E^*r$. If contacts are spread over the plane with a density η (per square micron), a ring distant r will contain $\eta.(2\pi\, r\, dr)$ contacts: and lower the height by

$$2\pi\eta(P/\pi E^*r)\, r\, dr = 2\eta(P/E^*)\, dr$$

But the load on the circle will be $\pi\, r^2 p_{nom}$: carried by $\eta\pi r^2$ contacts: so $P=p_{nom}/\eta$ and the ring of contacts will lower the height by $2\,(p_{nom}/E^*)dr$.
So the effect of the whole plane of contacts will be $2\,(p_{nom}/E^*)\int dr$….

large!... but uniform, and really just changing the datum

Its not easy to see why the effect of a ***single*** near-neighbour should matter



$\mu \equiv (8/3)\eta\sigma\sqrt{(2B\beta)} = 1$

Co-operative interaction can be very real, as this study of contact between a sphere and a rough surface showed.

A Hertzian pressure distribution develops as the load increases.

Greenwood & Tripp 1967

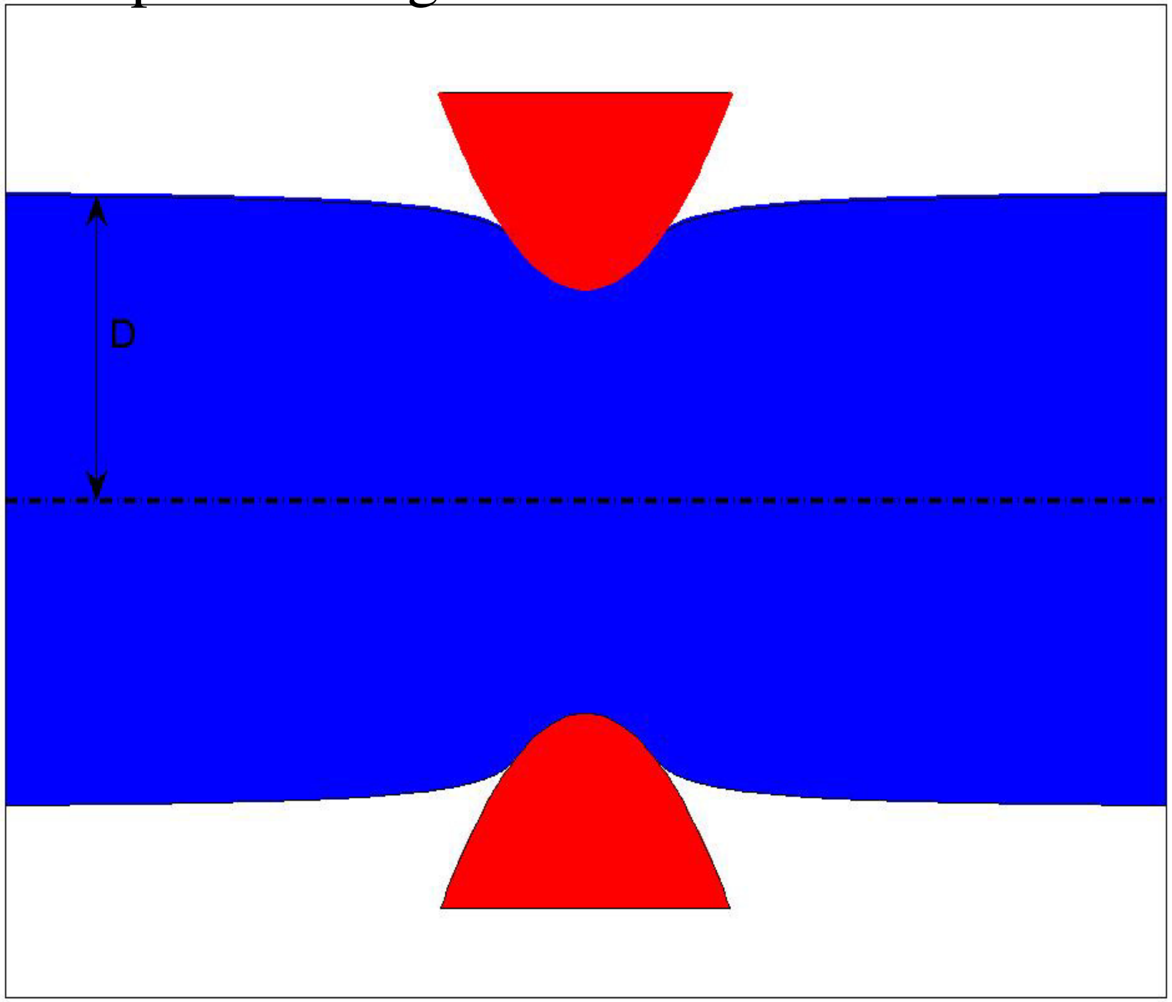
For contact between plane rough surfaces, we have to limit the contact area and assume a plane but finite indenter. Then we assume uniform pressure over this finite area…and use the known result that $\delta \approx p_{nom}\sqrt{A} / E^*$ to shift the datum.

Then the GW equation becomes

$$p_{nom} = \frac{4}{3} F_{3/2}(d/\sigma + p_{nom}\eta\sqrt{AR\sigma}\,)$$

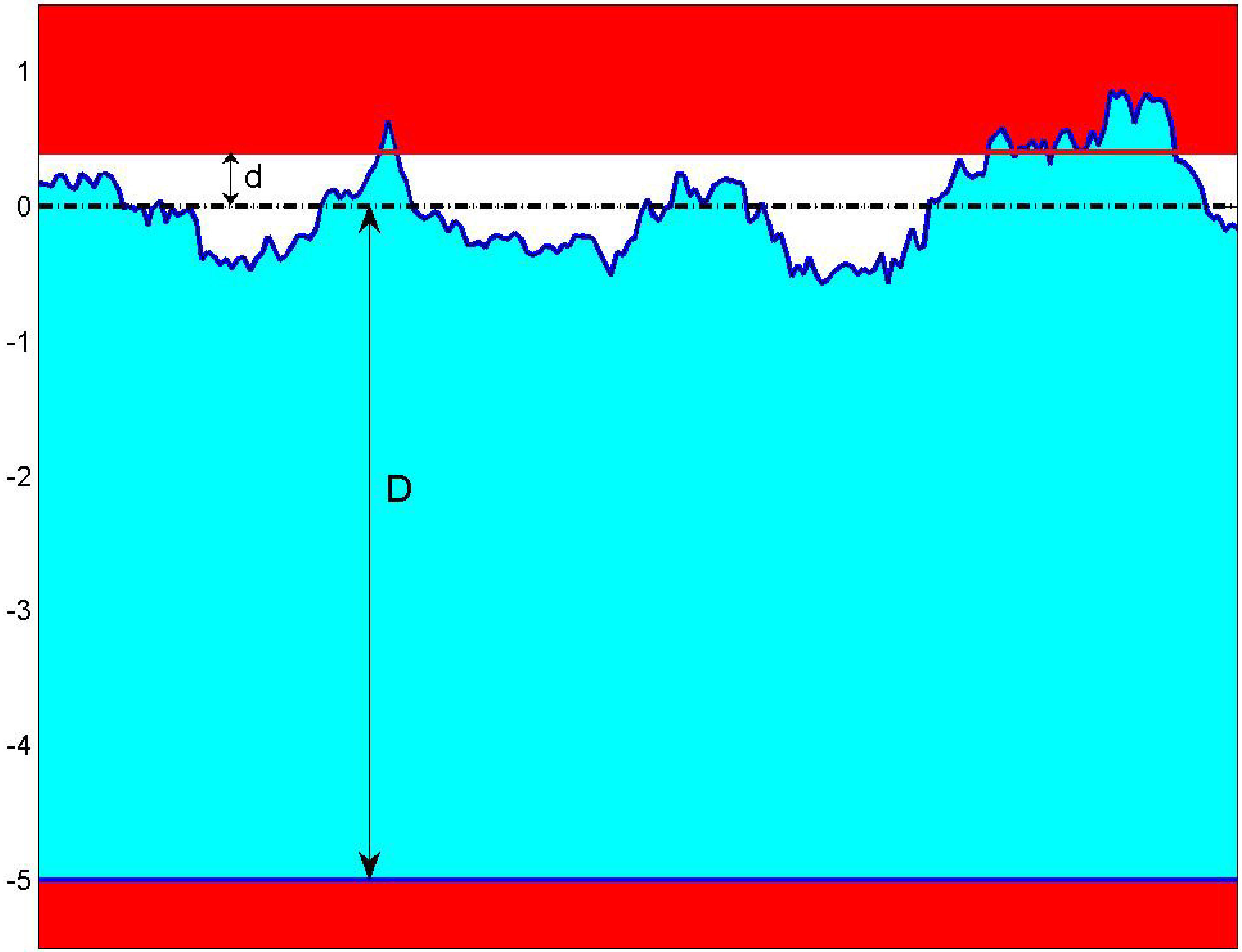
Ciavarella, Greenwood & Paggi 2008

The alternative to a finite indenter area is done in the numerical solutions do: put the roughness on a slab of finite thickness.

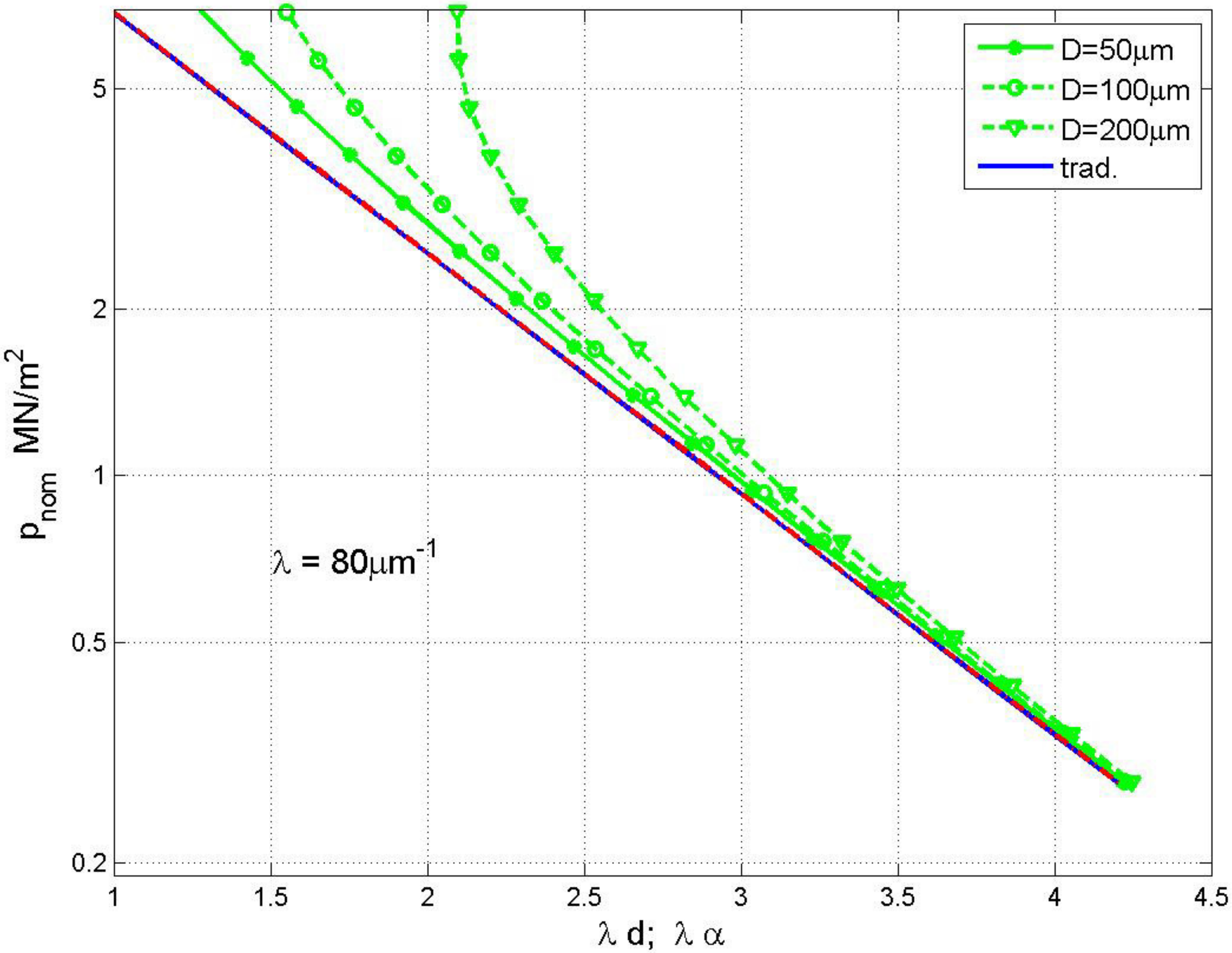


Hertz indentation of a finite slab

$$\delta = \left[\frac{9}{16} \frac{P^2}{RE'^2} \right]^{1/3} - \frac{P(1+\nu)(3-2\nu)}{2\pi E D}$$



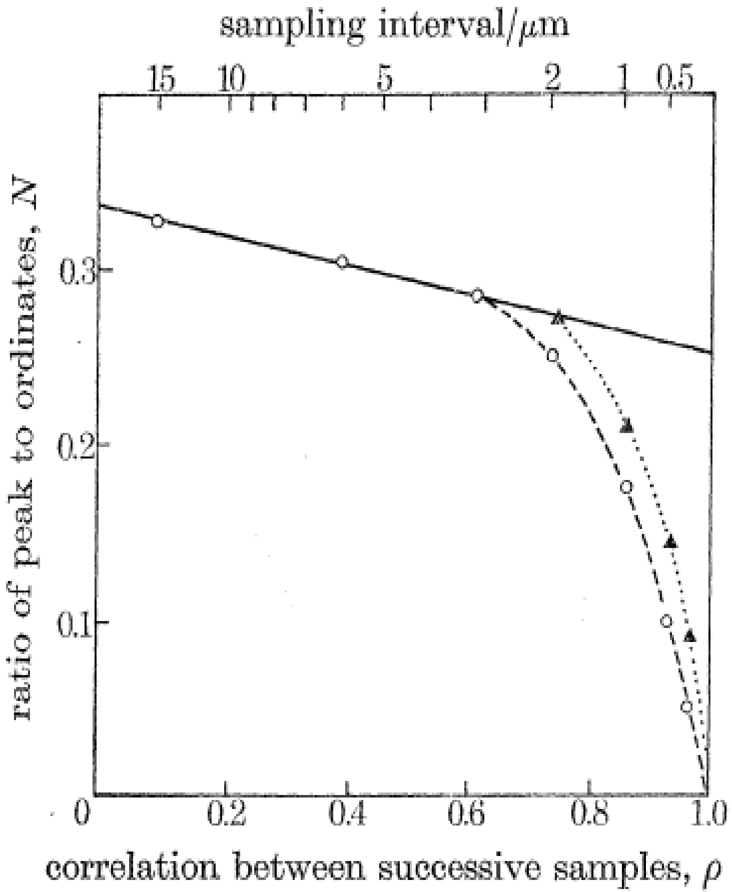
When a load is applied, both *d* ***and D*** reduce!

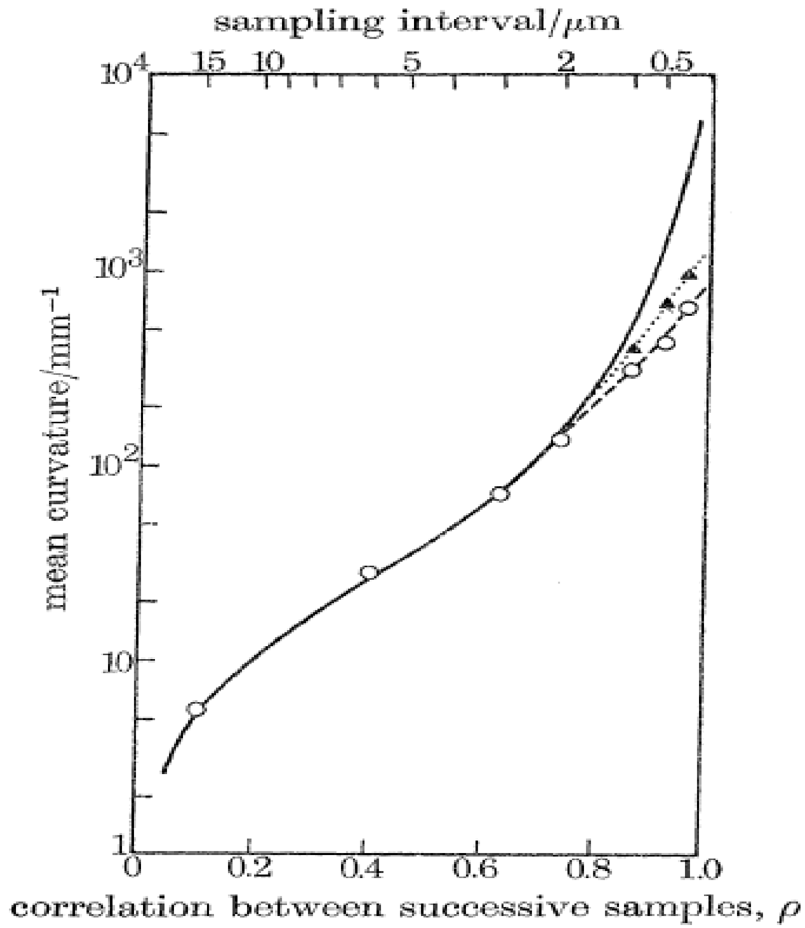


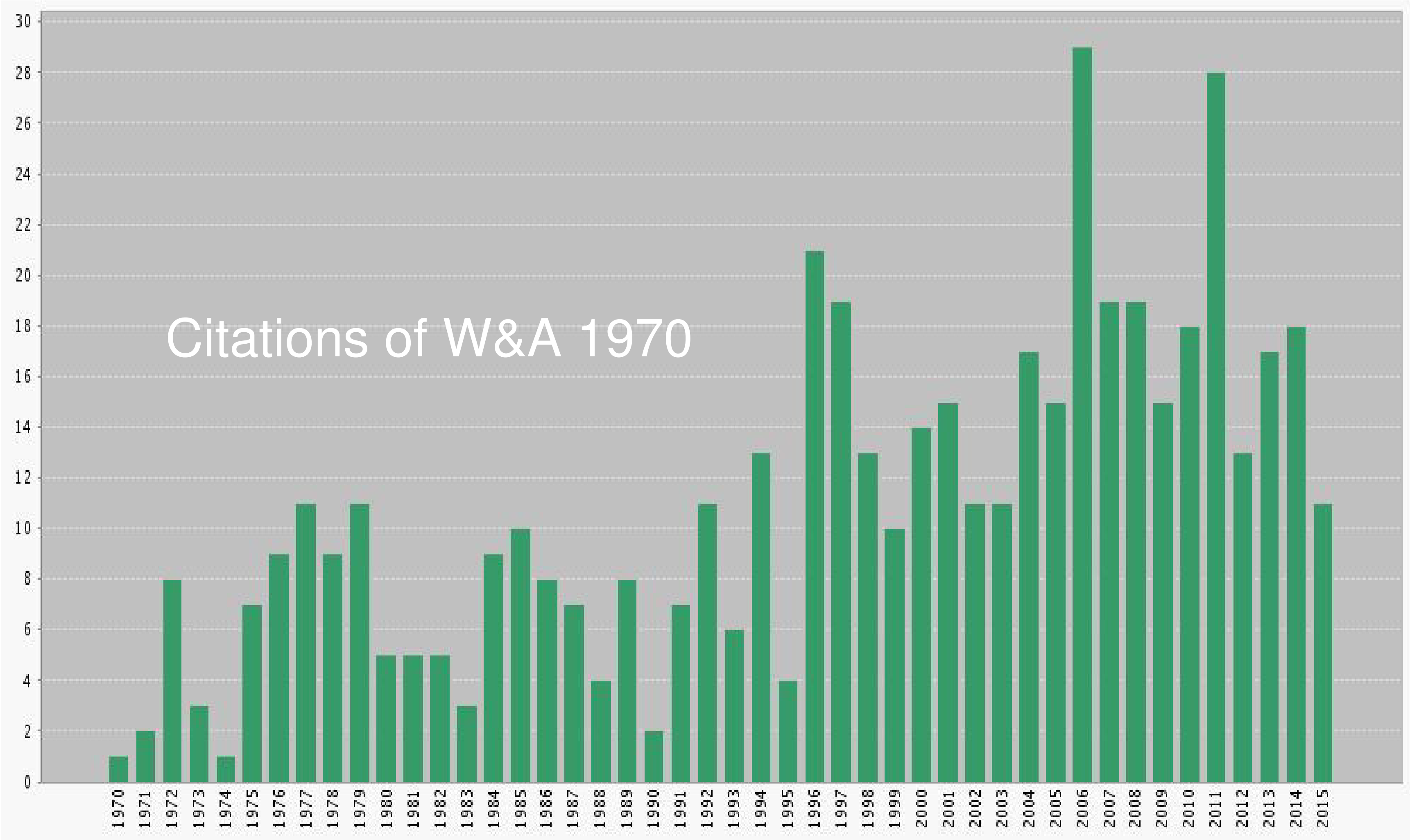
A finite slab thickness has a negligible effect on the separate Hertz Contacts: but the slab compression can dominate the behaviour

Whitehouse & Archard (1970) should have seen the end of GW !

For they showed it all depends on the sampling interval…with their ground surface, with an exponential autocorrelation function, between 1 in 3 and 1 in 4 of all points will be ”peaks” : and the peak curvature varies by a factor of 200







This is the paper we need to celebrate (in 5 years time). . . and it should have killed G&W dead. For W&A showed, both theoretically and experimentally, how everything depends on the sampling interval: and that all our *toys*: especially asperity density and summit curvature….and second moment m_2, and Nayak's α, can all be anything we like: their values ***meaningless***.

So why only 500 citations ??

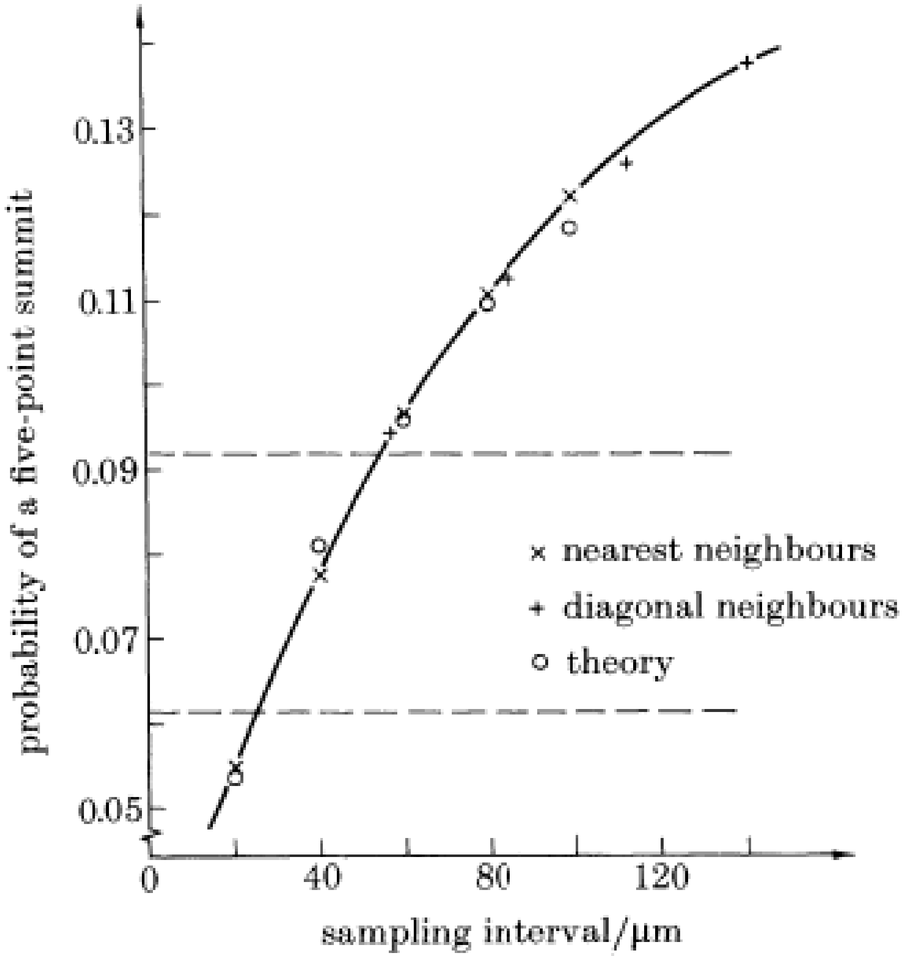


FIGURE 10. Experimental and theoretical summit densities for a grit-blasted surface (courtesy R. S. Sayles). The horizontal lines are the limits on the probability of true summits suggested by Sayles & Thomas.

Not a large variation in the summit density…until you realise this is the ***fraction*** of the sampled points which is a summit. As the sampling interval falls by a factor of 7, and the number of points / unit area increases by 50 ×, the summit density increased by 20 ×

So we can get nothing from measuring the surface in this way until we've chosen the relevant sampling interval.

So the spectral density approach is the better way..?

But the answers depend on m_2: the mean square profile slope.
And to find that, we measure the spectral density and integrate:
$m_2 = \int G(k)\, k^2\, dk$. Thus, for a power law $G(k) = A / k^p$, m_2 is infinite just as it was when we find it from the profile slopes.

We can make it finite by using a *finite sampling interval*: in the spectral density approach we do so by choosing an *arbitrary lower cut-off*.

Which do you prefer?

Ignore wavelengths shorter than $2\pi / k_1$: Sample at an interval Δ

? : ?

Only right that **Archard** should have the last word (as well as the first!).
He argued that one should worry only about the "main" structure, not about the "fine" structure.
And when you look at curve (b), would doing a G&W on those peaks (or the equivalent summits) be such a bad idea?

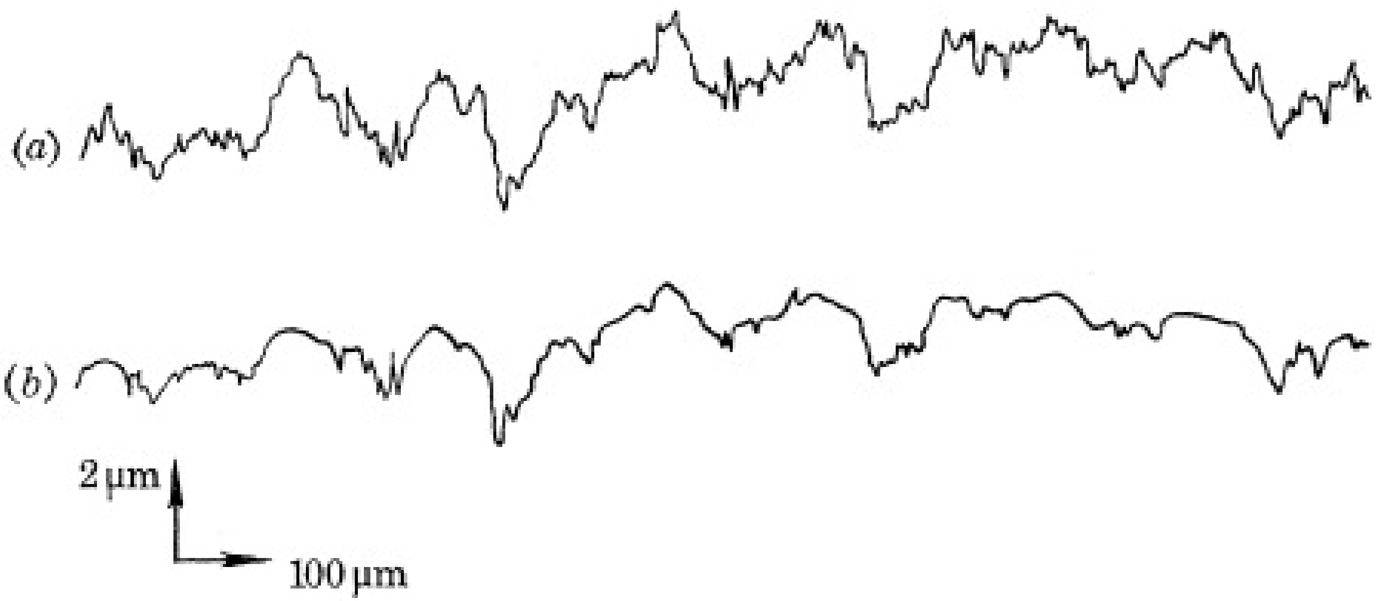


FIGURE 14. Talysurf profiles of cylindrical specimens used in lubricated friction experiments. (a) Original surface profile; (b) profile of the same line after one traversal of the load. Specimens 0.5 % C steel; 300 d.p.n.; 0.635 mm diameter; load 25 N.

Whitehouse & Archard (1970)

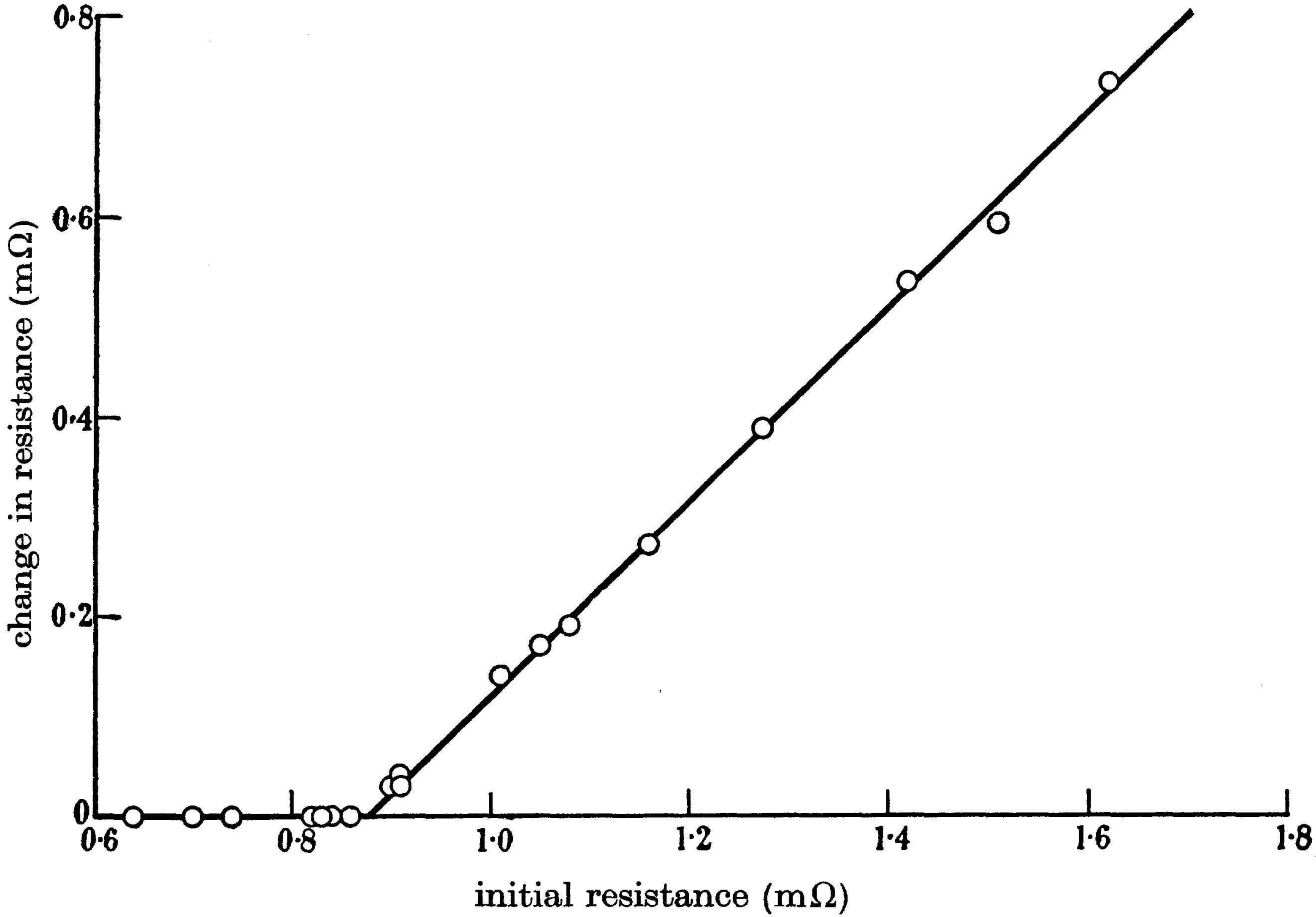


FIGURE 2. Effect of passing pulses of peak value 156 A through a series of constriction resistances.

Bowden & Williamson Proc. Roy. Soc. **246** (1958) p1

The Greenwood & Williamson Theory !

Kohlrausch (1901) found a simple relation between the voltage drop U across a contact and the maximum temperature rise θ_m:

$$U^2 = 8\int_0^{\theta_m} \lambda\rho \, d\theta$$

where λ is the thermal conductivity and ρ the electrical resistivity.

Using the Lorentz relation $\lambda\,\rho = LT$,

where L is the Lorentz constant 3.41×10^{-3}

the voltage drop needed to raise the contact temperature from 20^0C to 1063^0C will be 0.41 V.

But Williamson found changes began at 0.38 V

If equilibrium is possible, the Kohlrausch equation holds. But there is a second equation, relating the current to the temperature rise:

$$I \times \left(R_0 / \rho_0\right) = 2\int_0^{\theta_m} \left[2\int_{\theta}^{\theta_m} \lambda\rho \, d\theta\right]^{-1/2} \lambda \, d\theta$$

with an unexpected implication:

there is a maximum current which can be passed through the contact, unrelated to the melting point.

And quite probably, it corresponds to a temperature rise of $950^0 C$, and so, by Kohlrausch, to a voltage of 0.38 V

Why "quite probably" ?

Because we don't really have adequate data for gold at $900\,^0 C$ +

But we do have adequate data for iron

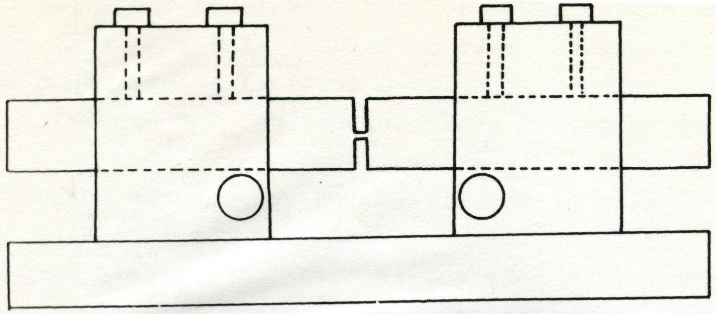


Fig.9a.

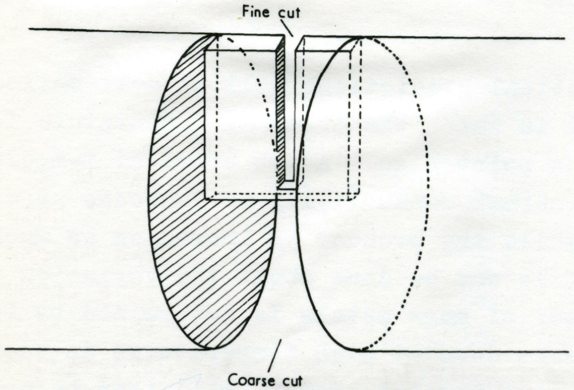


Fig.9b.

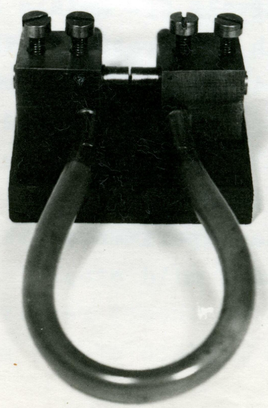


Fig.9c.

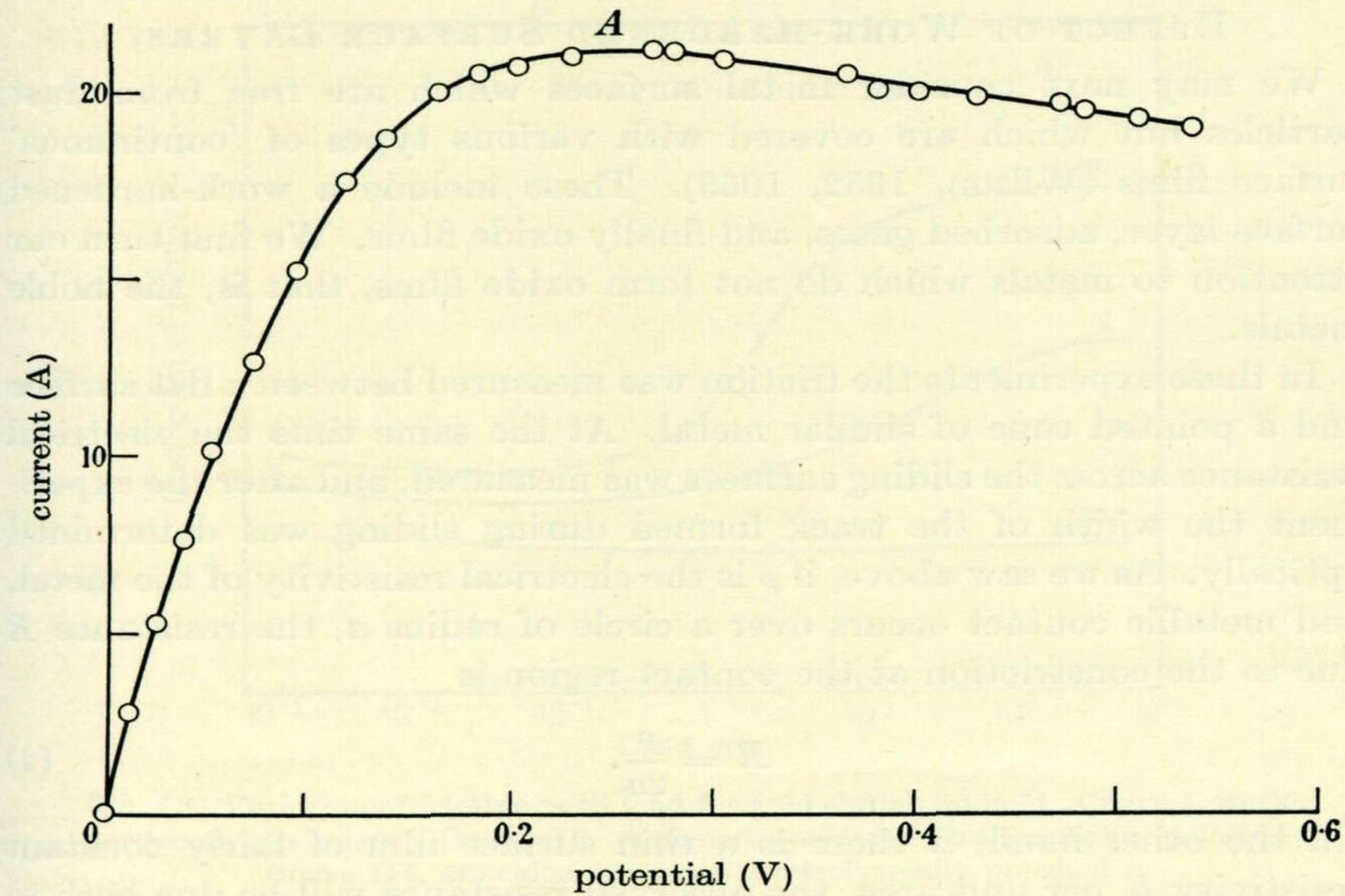
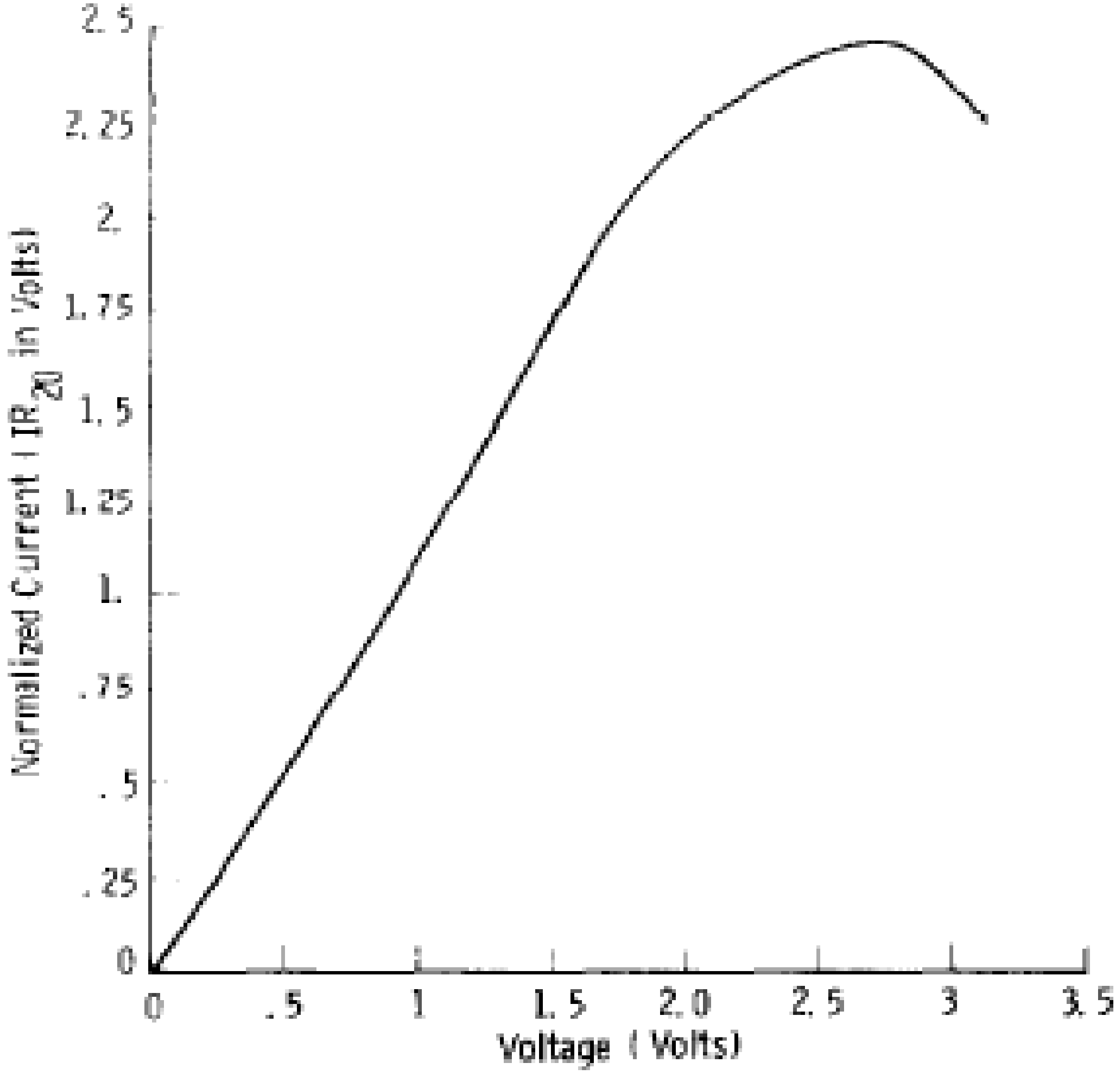


Fig. 17. Observed behaviour of Armco iron conductor in a low impedance circuit.

Greenwood & Williamson Proc. Roy. Soc. **246** (1958) p13



Relation between current and voltage for the contact between two pieces of ATJ graphite

Williamson & Allen Wear 78 (1982)

Dear Jim,

    I received the following from a Swiss scientist:

Dear Dr. Williamson
During my actual project I stumbled over a publication of yours quoted as No. (36) of chapter 1 in the book Electrical Contacts edited by Paul Slade:

Williamson, Greenwood: The constriction resistance between electroplated surfaces, Proc. of Int. Conf. on Electrical Contacts and Electromechanical Components, Appendix, Bejing, China, Oxford; Pergamon Press, 1989

We got the proceedings of that conference from the library of the ETH in Zürich but it doesn't contain your paper. The quotation includes "Appendix" which indicates that there is an additional volume but we couldn't find it so far. Is it possible to get a copy of that document from you directly? It would help us forward.

Thanks for your support in anticipation and best regards

Dr.-Ing. Hans Weichert

     I think this is the one where we considered the flow pattern in the plate and in the substrate separately, and adjusted them till the potentials at the plate/substrate interface matched.

     I can't find a copy: can you ?

Brian

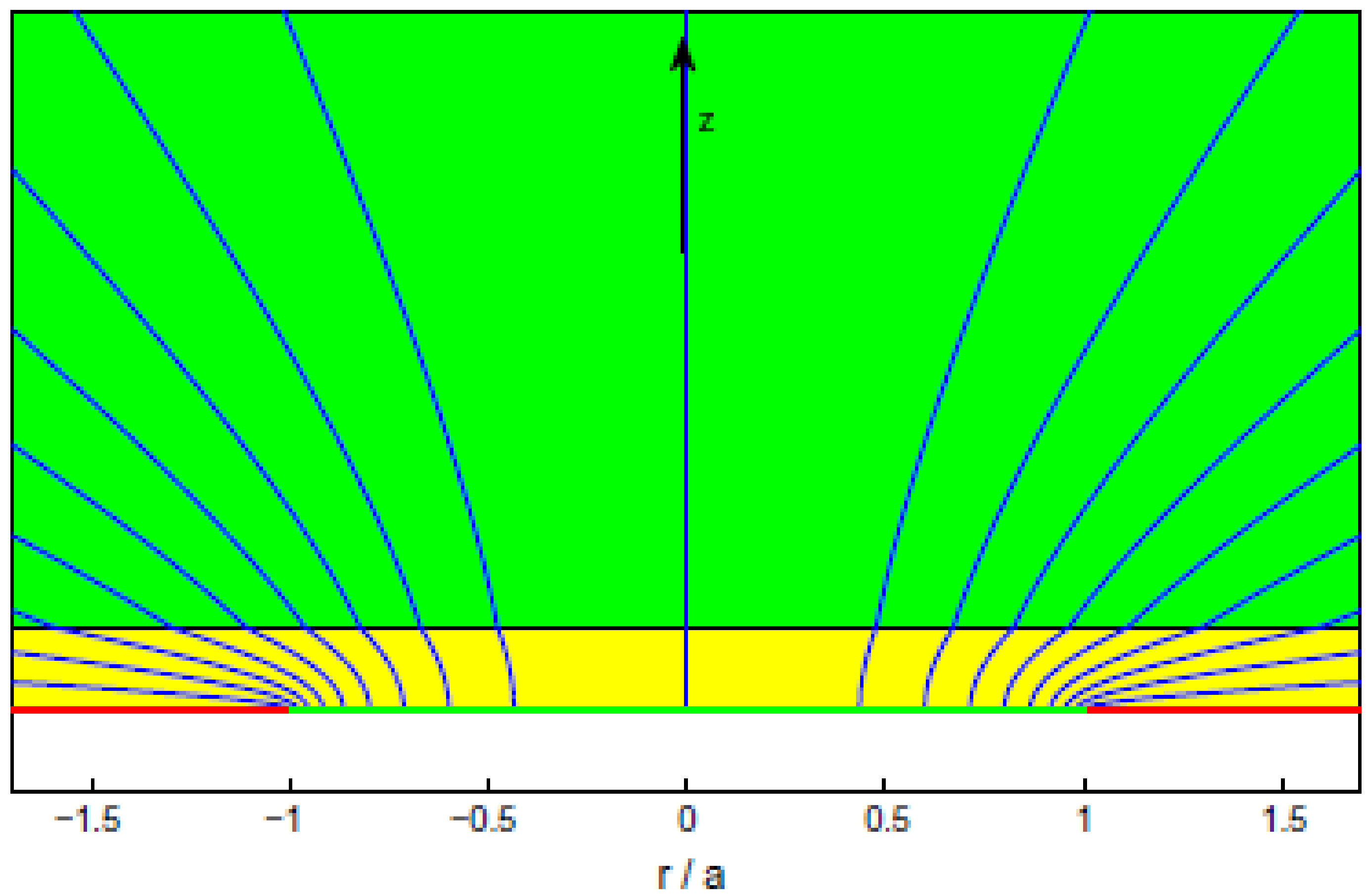


Fig. 1. Flux lines as heat enters a half-space through a circular contact area. [The flux lines will be kinked where the conductivity changes between plating and substrate].

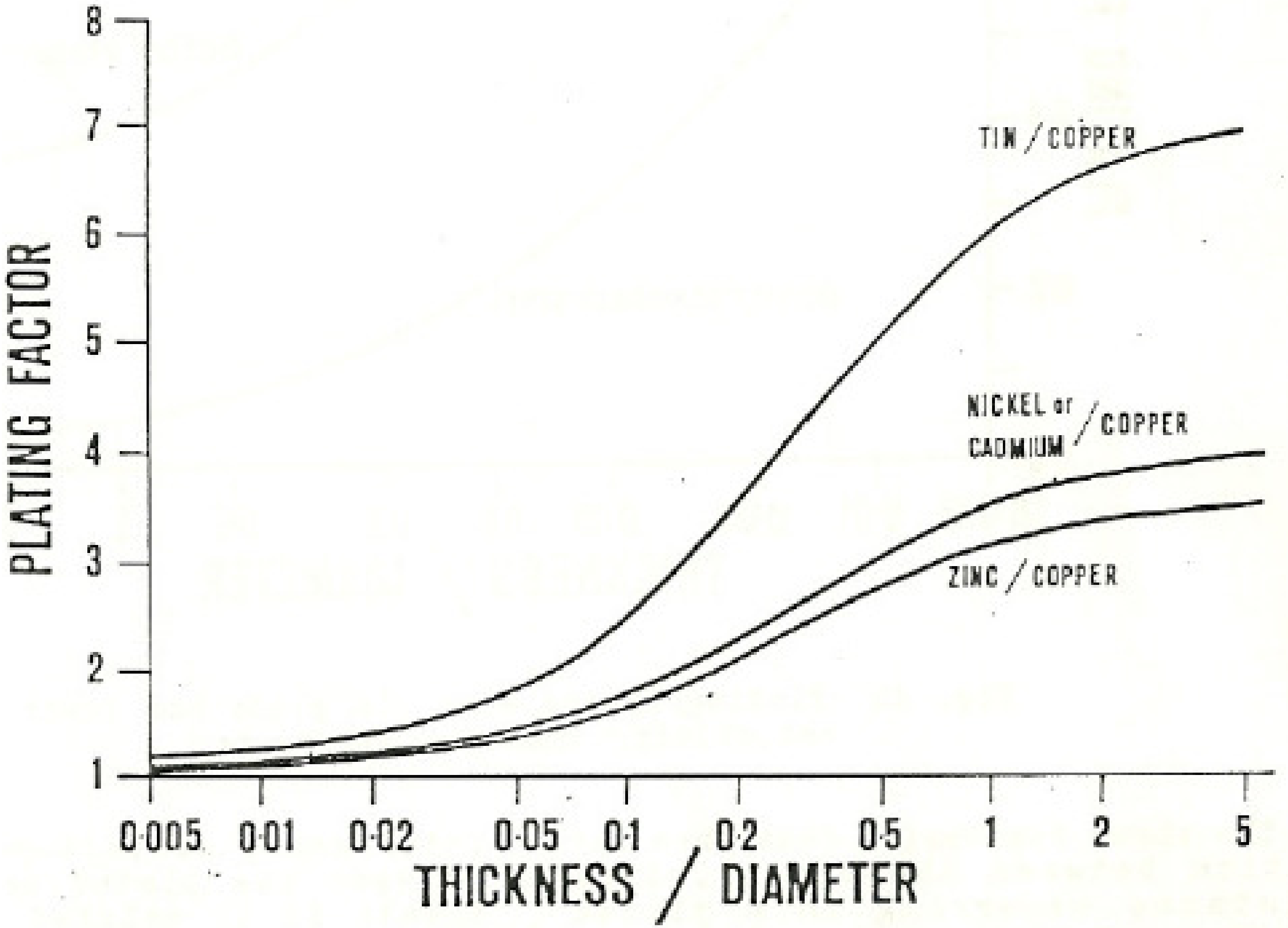
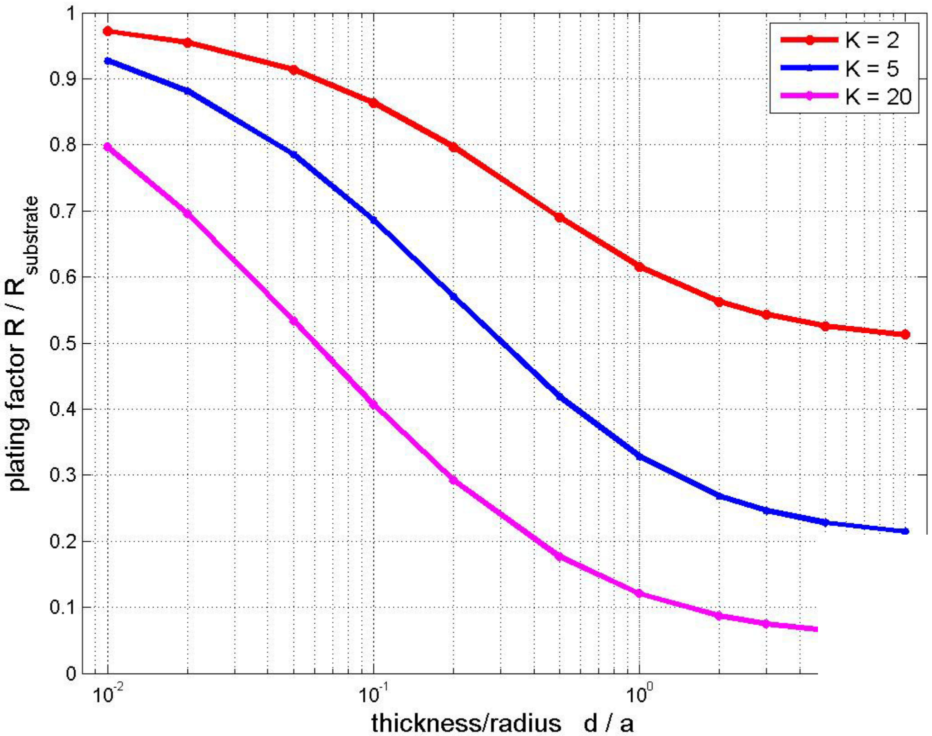


Fig. 4a Plating Factor when plate has higher resistivity than substrate

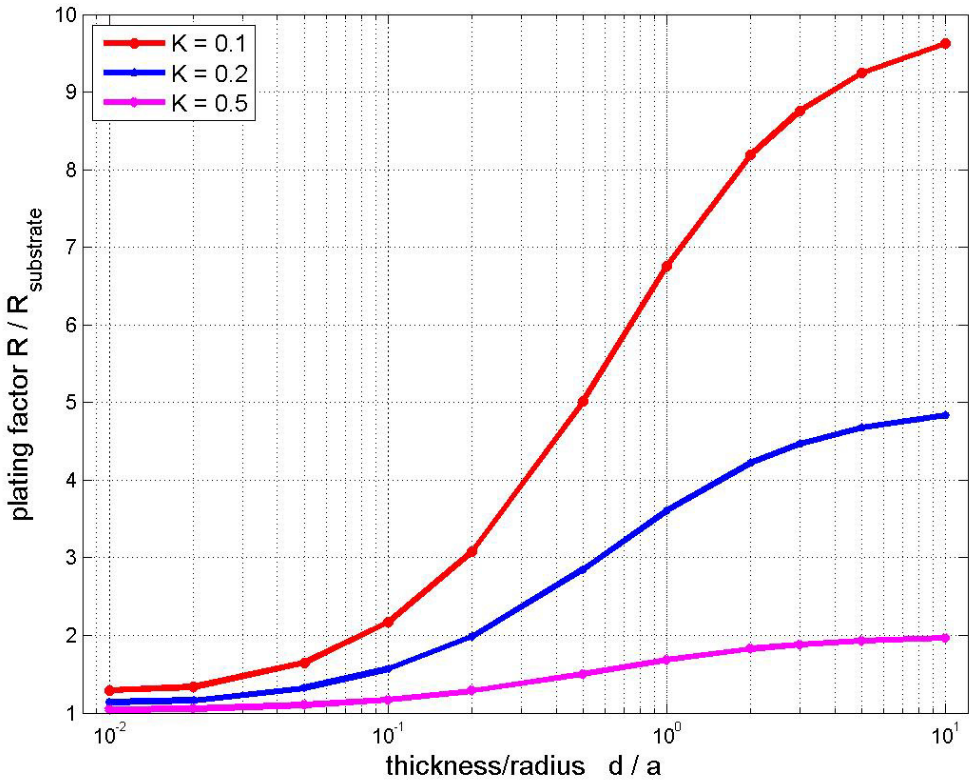
Williamson & Greenwood Beijing Electric Contacts conference 1989



$$R_{12} = \frac{4}{\pi}\int_0^{\infty} G(t\,(d/a))(\sin(t)/t)(J_1(t)/t)\,dt$$

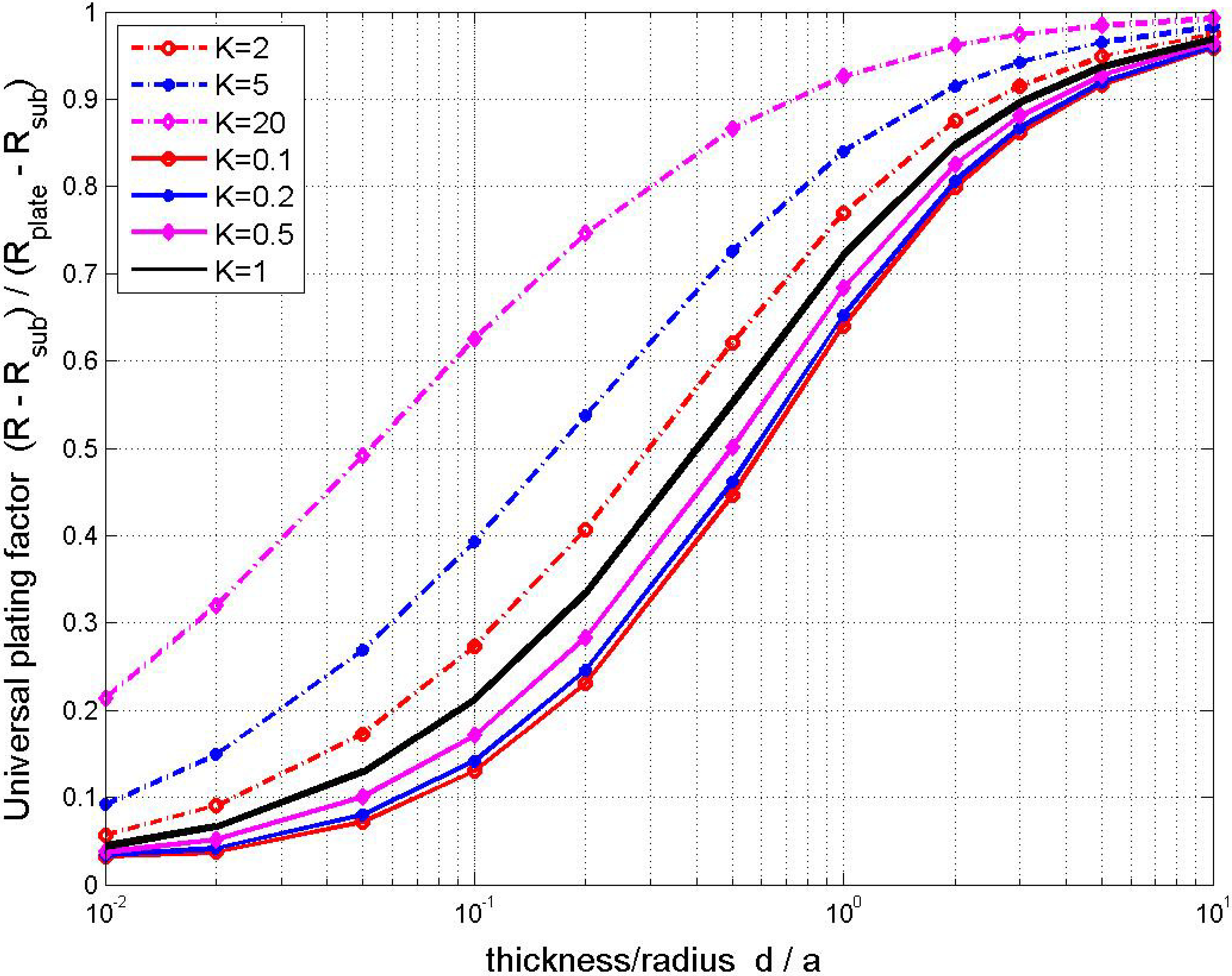
$$G(\lambda d) = \frac{\exp(\lambda d) + m}{\exp(\lambda d) - m}$$

$$m = \frac{\sigma_a - \sigma_b}{\sigma_a + \sigma_b}$$



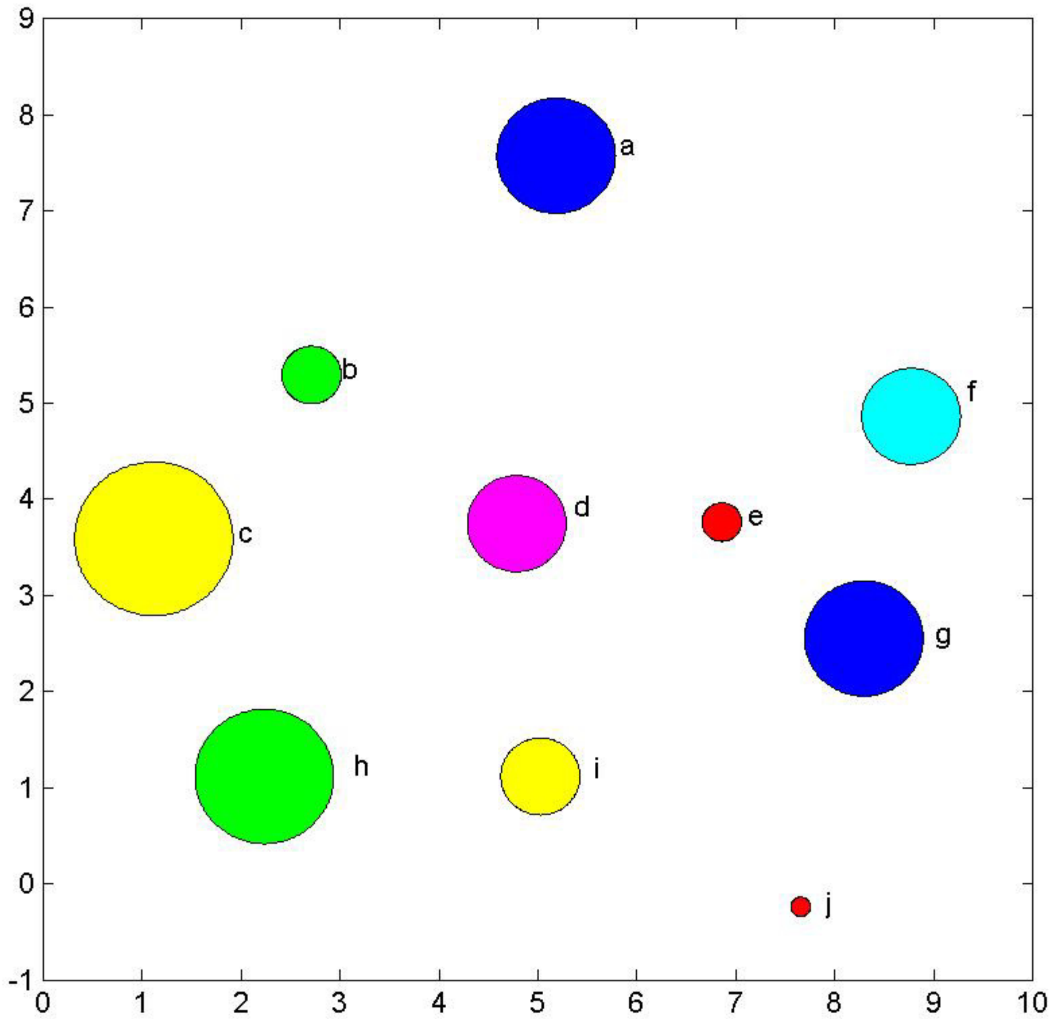
J A Greenwood

Int[l]. J. Heat Mass Transfer (2015) 90 p719

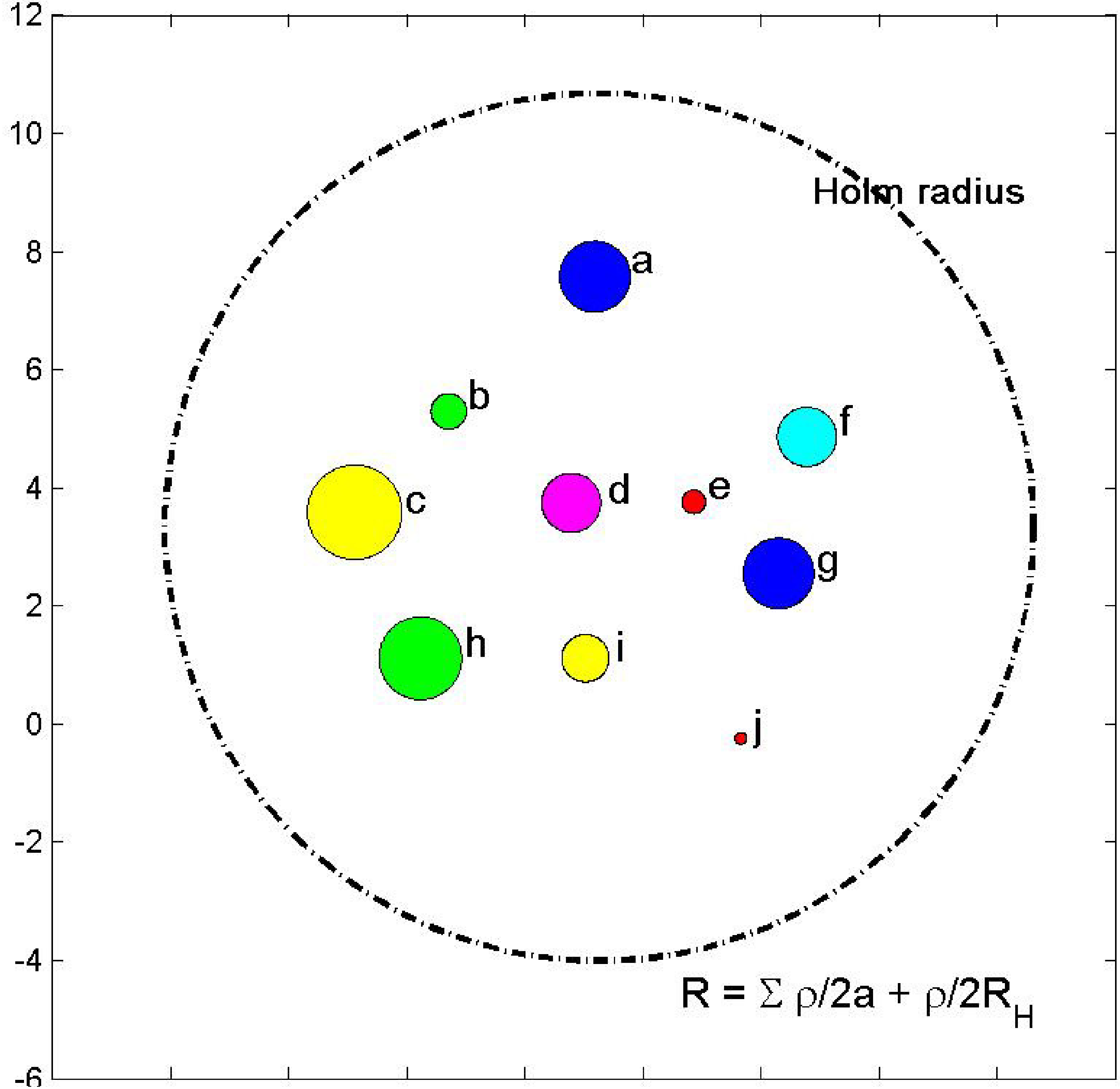

K=2
K=5
K=20
K=0.1
K=0.2
K=0.5
K=1
Universal plating factor (R - R_sub) / (R_plate - R_sub)
thickness/radius d / a
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
10^-2
10^-1
10^0
10^1

What can we say about the constriction resistance when we have a cluster, and not just a single contact ?

The contacts certainly do not act independently



Contact	a	b	c	d	e	f	g	h	i	j
Radius a	0.6	0.3	0.8	0.5	0.2	0.5	0.6	0.7	0.4	0.1
Current I	0.836	0.317	1.036	0.513	0.178	0.638	0.758	0.888	0.445	0.129
Ratio I/a	1.39	1.06	1.29	1.03	0.89	1.28	1.26	1.27	1.11	1.29



12
10
8
6
4
2
0
-2
-4
-6
Holm radius
a
b
c
d
e
f
g
h
i
j
R = Σ ρ/2a + ρ/2R_H

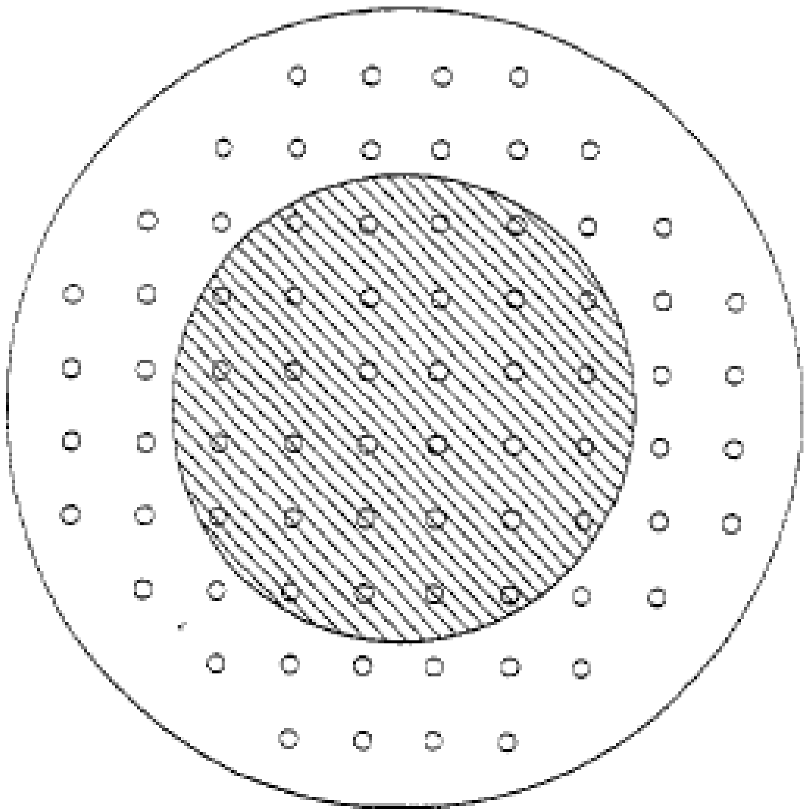
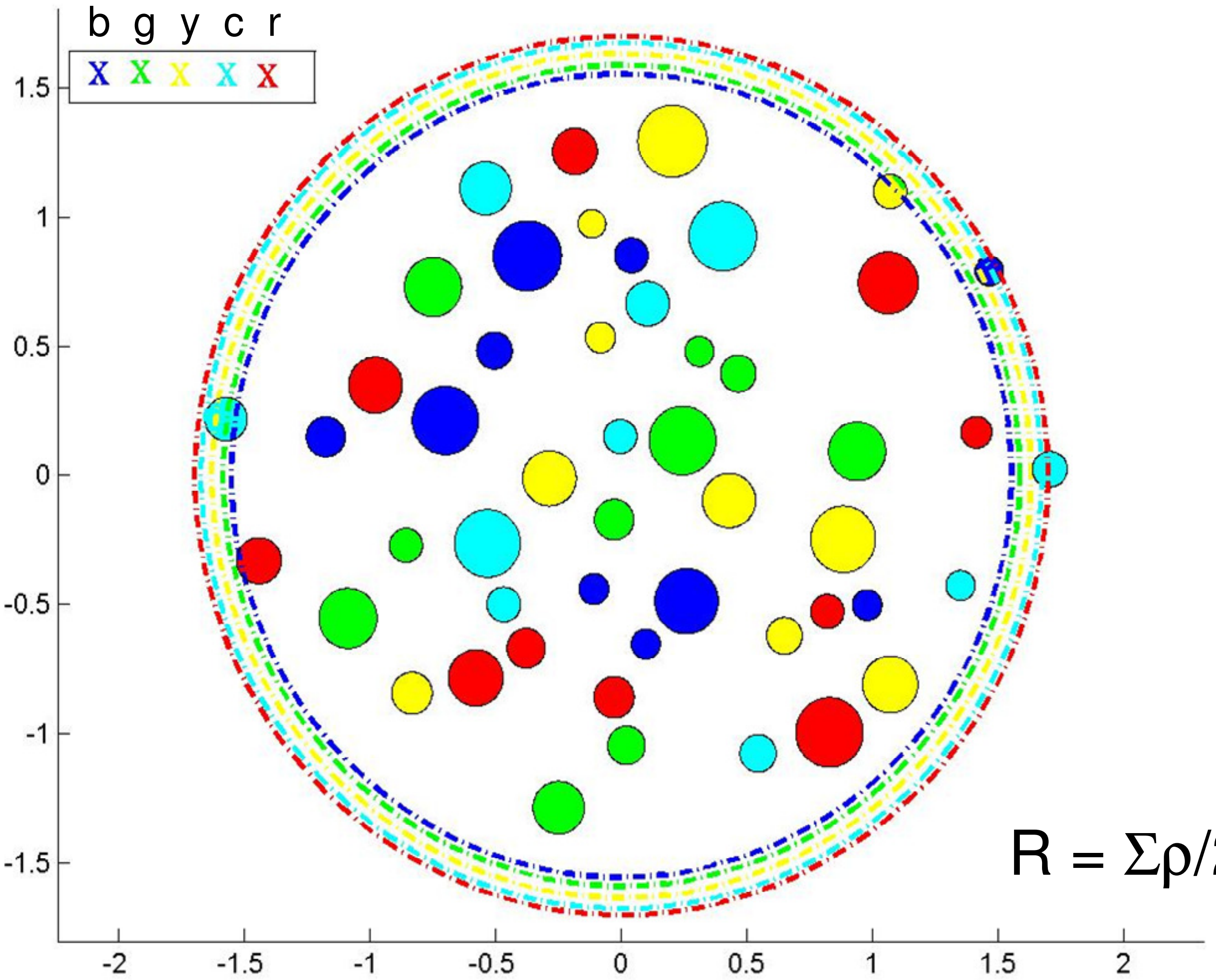


Figure 3. Resistance of a regular array of contact spots. The shaded area is the single contact with the same resistance; the outer circle has the Holm radius of the cluster.

$$R = \rho\left[\frac{1}{2n\bar{a}} + \frac{1}{2\alpha}\right] \quad \text{where} \quad \frac{1}{\alpha} = \frac{3\pi}{16n^2}\sum\sum\frac{1}{s_{ij}}$$

Greenwood Brit. J. Appl..Phys. (1966) **17** p1621

As the number of spots in a cluster increases,
the Holm circle remains very much the same size



$$R = \Sigma\rho/2a + \rho/2\alpha$$

References.

J A Greenwood & J B P Williamson
Contact of nominally flat rough surfaces.
[(2nd International Electric Contacts Conference, Graz, Austria (May 1964)]
Proc. Roy. Soc., **A295** p300-319 (1966)

J F Archard
Elastic deformation and the laws of friction
Proc. Roy. Soc., **A243** p190-205 (1958)

F. F. Ling
On Asperity Distributions of Metallic Surfaces
Journal of Applied Physics **29**, 1168 (1958); doi: 10.1063/1.1723395

T. E. Tallian , Y. P. Chiu , D. F. Huttenlocher , J. A. Kamenshine , L. B. Sibley & N. E. Sindlinger (1964)
Lubricant Films in Rolling Contact of Rough Surfaces,
A S L E Transactions, 7:2, 109-126; doi: 10.1080/05698196408972041

A. W. Bush, R.D. Gibson & T. R.Thomas
The Elastic Contact of a Rough Surface
Wear, **35** (1975) 87 - 111

J A Greenwood
A simplified elliptic model of rough surface contact.
Wear **261** (2006) p191-200
doi:10.1016/j.wear.2005.09.031

G Carbone
A slightly corrected Greenwood & Williamson model predicts asymptotic linearity between contact area and load.
Journal of the Mechanics and Physics of Solids **57** (2009) 1093-1102

D J Whitehouse & J F Archard
The properties of random surfaces of significance in their contact
Proc. Roy. Soc. **A316** p97-121 (1970)

J A Greenwood & J B P Williamson
Electrical conduction in solids. II Theory of temperature-dependent conductors.
Proc. Roy. Soc. **A246** p13-31 (1958)

J B P Williamson & J A Greenwood
Constriction resistance between electroplated surfaces.
Proc. Intl. Conference on Electrical Contacts….Beijing, May 1989 p563-568

J A Greenwood
Constriction resistance of a plated solid
International Journal of Heat and Mass Transfer **90** (2015) 719–727
http://dx.doi.org/10.1016/j.ijheatmasstransfer.2015.03.032

J A Greenwood
Constriction resistance and the real area of contact.
[Holm Electric Contacts Conference, Orono, Maine. ?July 1965?]
Brit. J. Appl. Phys., **17** p1621-1632 (1966)